JOINT MODELS FOR LONGITUDINAL AND SURVIVAL DATA

Lili Yang

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biostatistics,
Indiana University

December 2013

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Sujuan Gao, Ph.D., Chair

Menggang Yu, Ph.D.

Doctoral Committee

Wanzhu Tu, Ph.D.

September 23, 2013

Christopher M. Callahan, M.D.

Terrell Zollinger, Ph.D.

DEDICATION

To My Family

ACKNOWLEDGMENTS

I would like to express sincere gratitude to my advisor Dr. Sujuan Gao for her constant guidance, encouragement and support in my study. I particularly appreciate Dr. Menggang Yu for his guidance and suggestion on my dissertation. I also thank the other committee members Dr. Wanzhu Tu, Dr. Christopher M. Callahan and Dr. Terrell Zollinger for their insights and comments on my dissertation.

It has been a pleasant experience to study in this department, and in this university. I would like to convey my gratitude to all the people who have made the resources available to me, without which I would not be able to complete my study.

Finally, I would like to thank my husband Xiao Ni and family, for their unconditional love, encouragement and support.

Lili Yang

JOINT MODELS FOR LONGITUDINAL AND SURVIVAL DATA

Epidemiologic and clinical studies routinely collect longitudinal measures of multiple outcomes. These longitudinal outcomes can be used to establish the temporal order of relevant biological processes and their association with the onset of clinical symptoms. In the first part of this thesis, we proposed to use bivariate change point models for two longitudinal outcomes with a focus on estimating the correlation between the two change points. We adopted a Bayesian approach for parameter estimation and inference. In the second part, we considered the situation when time-to-event outcome is also collected along with multiple longitudinal biomarkers measured until the occurrence of the event or censoring. Joint models for longitudinal and time-to-event data can be used to estimate the association between the characteristics of the longitudinal measures over time and survival time. We developed a maximum-likelihood method to joint model multiple longitudinal biomarkers and a time-to-event outcome. In addition, we focused on predicting conditional survival probabilities and evaluating the predictive accuracy of multiple longitudinal biomarkers in the joint modeling framework. We assessed the performance of the proposed methods in simulation studies and applied the new methods to data sets from two cohort studies.

Sujuan Gao, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In this thesis research, several topics related to joint models for longitudinal and survival data analysis were investigated. Longitudinal data analysis has been widely applied to a single longitudinal outcome in various medical research areas, including basic science research, clinical trials and epidemiological studies. In practice, however, many studies often collect multiple longitudinal outcomes and joint models can be used to address interesting scientific questions regarding the relationships among these multiple processes. In addition, often times, a time-to-event outcome is also collected along with multiple longitudinal outcomes in medical research studies. Joint models for multiple longitudinal outcomes and time-to-event data can be used to assess the association between the time-to-event outcome and multiple longitudinal outcomes.

We developed several novel approaches for analyzing multiple longitudinal outcomes, and multiple longitudinal outcomes with time-to-event data. First, we introduce bivariate random change point models for joint modeling of bivariate longitudinal outcomes. Second, we present joint models for multiple longitudinal outcomes and time-to-event data. Finally, we focus on predicting conditional survival probabilities and evaluating the improved predictive ability by adding new longitudinal biomarkers in the joint models.

1.1 Bivariate Random Change Point Models for Longitudinal Outcomes

In most longitudinal analysis a single longitudinal outcome, measured repeatedly over time, was the focus of investigation on identifying the longitudinal trend or factors associated with longitudinal change. For example, in longitudinal cohort studies of dementia, cognitive

function, activities of daily living (ADL), and physiological measures such as blood pressure (BP), height and weight are collected repeatedly from participants over a relatively long follow-up period. Many of these functional measures are assumed to be relatively stable across the life span and may start to decline or increase with the onset of underlying diseases. The time point when an individual start the decline is called a change point. It is therefore of interest to determine the change point when individual declines on a specific outcome. Furthermore, it is perhaps more interesting to determine whether the change point of one longitudinal measure is associated with the change point of another longitudinal measure, thus offering potential evidence of a temporal association linking two or more biological processes. Our motivating example data for joint modeling of bivariate longitudinal data came from dementia studies. It is well known that subjects with dementia or cognitive impairment suffer weight loss, which was often attributed to the fact that these subjects often forget to eat resulting in nutritional deficit. However, Buchman et al. (2005) also reported that weight loss precedes dementia diagnosis. Thus, it is of scientific interest to examine the temporal relationship between these two outcomes to determine whether cognitive decline leads to weight loss or whether weight change precedes cognitive impairment. In this research, we developed several bivariate random change point models for two longitudinal outcomes with a particular focus on the correlation between the change points of the two trajectories.

1.2 Joint Models for Multiple Longitudinal Processes and Time-to-event Outcome

In both epidemiological and clinical trial studies, the time-to-event outcome is often collected along with multiple longitudinal biomarkers that are repeatedly measured until the occurrence of the event or censoring. There are two general joint modeling strategies with

different model interpretations discussed (Little, 1993; Little and Rubin, 2001). The pattern-mixture model is used when the primary interest is inference of the longitudinal process with the time-to-event outcome considered a missing data phenomenon. On the other hand, if the primary inference is the time-to-event outcome and to determine whether the longitudinal outcome is associated with the event process, selection models are more appropriate. Sousa (2011) gave a brief introduction to the two types of joint modeling frameworks. In the second topic of this dissertation research, we focus on the latter type of joint model - the random selection model, which could be formulated as $[Y, F, U] = [U][Y|U][F|Y]$, where Y is the longitudinal measures, F is the time-to-event outcome, and U presents the random effects.

Traditional survival models have typically characterized exposures by a single measure at study baseline or as an average over a relatively short period of time. Such exposure characterization fails to capture any changes or variability over the potentially long latency period prior to an event. An extension to this standard survival model is the Cox model involving time-dependent covariates using counting process formulation and partial likelihood theory (Andersen et al., 1993; Andersen and Gill, 1982; Fleming and Harrington, 1991). However, this model has a strong assumption that the time-dependent covariates are measured without error and precisely predictable. In practice, this is not realistic because often times the longitudinal biomarker measures are not observed at the event or censoring time point. Therefore, an ideal model should not only allow the examination on the contribution from various attributes of the longitudinal outcomes in order to establish the association between the longitudinal outcome and the time to event but also takes the measurement errors of the longitudinal biomarkers into account. With this strong motivation, a framework of joint models of longitudinal and survival data was proposed (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997). We focused on this type of joint

models with multiple longitudinal processes and a time-to-event outcome, and developed a maximum-likelihood method for the parameter estimation.

1.3 Dynamic Predictions in Joint Models for Multiple Longitudinal Processes and Time-to-event Outcome

In the second topic, we proposed a maximum-likelihood method for parameter estimation of joint models for multiple longitudinal biomarkers and a time-to-event outcome, where the main interest is to assess the associations between the multiple longitudinal biomarkers and the risk of an event. The estimated associations can help researchers better understand the relationship between the time-to-event outcome and multiple longitudinal biomarkers. However, in reality, it may be more clinically relevant to study how well the longitudinal biomarkers predict the event risk. In this work, we concentrated on predicting conditional survival probabilities and assessing the predictive accuracy of the joint models of multiple longitudinal biomarkers and a time-to-event outcome. In particular, we used traditional criterion, the area under the receiver operating characteristic (ROC) curve (AUC) (Hanley and McNeil, 1982), to evaluate the predictive accuracy. However, several studies have demonstrated that AUC is not sensitive and appropriate in evaluating the improvement in predictive ability by adding new biomarkers in the model (Cook, 2007; Harrell, 2001; Janes et al., 2008; Moons and Harrell, 2003). In the past few years novel predictive criteria have been proposed for binary and time-to-event outcomes, including the above average risk difference(AARD) and the mean risk difference(MRD) (Pepe et al., 2008; Pepe and Janes, 2012). We applied AARD and MRD to the joint modeling framework and evaluated their ability in quantifying the improved prediction by adding new longitudinal biomarkers.

Chapter 2

Bivariate Random Change Point Models for Longitudinal Outcomes

2.1 Abstract

Epidemiologic and clinical studies routinely collect longitudinal measures of multiple outcomes, including biomarker measures, cognitive functions, and clinical symptoms. These longitudinal outcomes can be used to establish the temporal order of relevant biological processes and their association with the onset of clinical symptoms. Univariate change point models have been used to model various clinical endpoints, such as CD4 count in studying the progression of HIV infection and cognitive function in the elderly. We proposed to use bivariate change point models for two longitudinal outcomes with a focus on the correlation between the two change points. Three types of change point models are considered in the bivariate model setting: the broken-stick model, the Bacon-Watts model and the smooth polynomial model. We adopted a Bayesian approach using a Markov chain Monte Carlo sampling method for parameter estimation and inference. We assessed the proposed methods in simulation studies and demonstrated the methodology using data from a longitudinal study of dementia.

2.2 Introduction

Longitudinal epidemiologic and clinical studies routinely collect repeated measures of multiple outcomes. For example, in longitudinal studies of dementia, cognitive function measures, activities of daily living (ADL) measures, physical function measures such as height and weight, neurological measures, and psychosocial measures are collected repeatedly from par-

ticipants over a relatively long follow-up period. In recent years, research on Alzheimer's disease (AD) has come to the consensus that both AD pathological processes and the clinical decline occur gradually, with dementia at the end stage of many years of accumulation of these pathological changes (Jack et al., 2010). An additional feature of AD is that biological changes begin to develop decades before the presentation of earliest clinical symptoms. Longitudinal measures of biomarkers, cognitive functions, and clinical symptoms will enable researchers to establish the temporal order of relevant biological processes and their association with the onset of clinical symptoms.

Change point models are useful as an alternative to linear models to determine when changes have taken place in an event window. Change point models with one change point and two linear phases are most commonly used, because many biological mechanisms can be readily modeled. To account for individual variability, random change point models have been further formulated by including flexible subject-specific random effects to capture both population trends and individual-level variations. Univariate change point models have been used to model various clinical endpoints such as CD4 count in studying the progression of HIV infection and AIDS (Ghosh and Vaida, 2007; Kiuchi et al., 1995; Lange et al., 1992) and cognitive function in studying dementia in the elderly (Dominicus et al., 2008; Hall et al., 2003; Jacqmin-Gadda et al., 2006; van den Hout et al., 2010).

The simple change point model with an abrupt transition is referred to as the broken-stick model (Dominicus et al., 2008; Ghosh and Vaida, 2007; Kiuchi et al., 1995), which has the advantage of detecting a significant departure in direction and volatility from the immediate past. However, the broken-stick model is not always appropriate in practice because a sudden change in direction may not be realistic. The non-continuity at the change point of the broken-stick model may also cause numerical issues in parameter estimation.

Two types of smooth change point models were proposed by van den Hout et al. (2010): the Bacon-Watts model (Bacon and Watts, 1971) and a smooth polynomial model.

There have been a few studies on the joint modeling of bivariate random change point model for longitudinal outcomes. Hall et al. (2001) simultaneously estimated two different change points of two longitudinal measures of cognitive function. Jacqmin-Gadda et al. (2006) constructed joint models between a random change point model for a longitudinal outcome and a lognormal model for time-to-event data. In this paper, we consider bivariate change point models for two longitudinal outcomes with a focus on the correlations between the two change points. Motivated by data from a longitudinal study of dementia, we developed joint models for bivariate longitudinal outcomes under the aforementioned modeling frameworks: the random broken-stick model, the random Bacon-Watts model and the random smooth polynomial model. The proposed bivariate change point models take the correlation structure into account and provide a useful framework to assess the correlation between the two change points and their temporal order. The proposed methodology is applicable to other studies in which determining the order of biomarker changes is needed. We adopted a Bayesian estimation approach using Markov chain Monte Carlo (MCMC) for a computational and inferential framework for the bivariate random change point models. We assessed the performance of the proposed method in simulation studies and demonstrated the methodology using data from a longitudinal study of dementia.

The remainder of this chapter is organized as follows. Section 2.3 describes a longitudinal study of dementia as a motivating example. In Section 2.4, we present three bivariate random change point models, the Bayesian methodology for parameter estimation, and statistical inference. A series of simulation studies were carried out to compare the performances of the three joint models and results are presented in Section 2.5. In Section 2.6

we apply the proposed methods to the example data set. We conclude with a discussion in Section 2.7.

2.3 The Indianapolis-Ibadan Dementia Study

The Indianapolis-Ibadan Dementia Study (IIDS) is a longitudinal comparative epidemiology study designed to investigate risk factors associated with dementia and AD. The study enrolled and maintained two cohorts of elderly participants, one consisting of African Americans living in Indianapolis, Indiana, and the other consisting of Nigerians living in Ibadan, Nigeria. Details about the study have been published (Hendrie et al., 2001, 1995). The data used for the current paper come from the Indianapolis cohort. Briefly, 2212 African-American adults aged 65 and older living in Indianapolis were enrolled in the study in 1992. The study participants were followed for up to 17 years and underwent regularly scheduled cognitive assessments and clinical evaluations approximately every2 or 3 years. In this ongoing study, there were 7 evaluations by the end of 2009.

The cognitive function of study participants was measured by the Community Screening Interview for Dementia (CSID), at baseline and at years $3, 6, 9, 12, 15$, and 17 with respect to the baseline. The CSID questionnaire (Hall et al., 1996) has been widely used as a screening tool for dementia. It evaluates multiple cognitive domains including language, attention, memory, orientation, praxis, comprehension, and motor response. For this analysis, we use a CSID score that incorporated all cognitive items from the screening exam some of which had not been utilized previously (Hall et al., 1996). The additional cognitive score items in the CSID include the East Boston story (immediate and delayed recall), 3 mental calculation items, the name of the state, the name of the president, and the name of the governor. In addition, unit weighting was used for object repetition, object recall, instruction command, and animal naming, with the exception that animal naming is capped at a maximum raw

score of 23 (95th percentile). The CSID total score ranges from 0 to 80, with higher scores indicating better cognitive function.

Also, height and weight measures from all participants were collected at each evaluation starting from year 3. Because obesity is associated with increased risk for diabetes, hypertension and cardiovascular diseases, conditions related to increased risk of dementia, it is therefore important to monitor weight change in this elderly cohort. It is widely known that subjects with dementia and cognitive impairment suffer weight loss, which can be attributed to the fact that these subjects often forget to eat. However, there are also reports that weight loss precedes dementia diagnosis (Buchman et al., 2005). In particular, in this cohort, we found that accelerated weight loss was associated with dementia or mild cognitive impairment (MCI) as early as 6 years prior to clinical diagnosis, supporting the hypothesis that weight loss is an early marker for the manifestation of the dementia disorder, including the early stage of MCI (Gao et al., 2011). It is important to examine the longitudinal trajectories of both cognitive function and weight measures to determine whether cognitive decline leads to weight loss or whether weight change proceeds cognitive impairment. Because both body weight and cognitive function are assumed to be stable over time and sudden changes may indicate underlying disease processes, we propose to use bivariate change point models to model cognitive trajectories and changes in body mass index (BMI) over time, with a particular focus on the correlation between the change points of the two trajectories. Here, BMI is defined as weight in kilograms divided by height in meters squared. We choose to use only two change points based on the study design of the IIDS data. The IIDS followed normal subjects without dementia to dementia diagnosis and no data were collected once a subject was diagnosed with dementia. Because both body weight and cognitive function in elderly subjects without dementia are assumed to be stable over time and sudden changes may indicate underlying dementia progression, we believe one

change point for each longitudinal trajectory should capture the decline in the pre-dementia or earlier dementia stage. It is possible that there exists a second change point, reflecting a rapid deterioration in both body weight and cognition just prior to death. However, because the IIDS did not conduct any follow-up evaluations in the subjects with dementia and our evaluation interval window of every 2 to 3 years may be too wide to capture the rapid changes in the second change points, we focused on models with only one change point.

Out of the 2212 IIDS participants enrolled at baseline, 441 had at least 5 cognitive measurements, of which 238 also had at least five BMI measurements. For modeling purposes, we restrict the data to participants with at least 5 measurements for both of cognitive function and BMI $N = 238$. Out of the 238 subjects with age ranges from 64.3 to 84.6 at baseline, 190 (79.8%) subjects were female. The mean baseline age was 70.4 (SD=4.8) years old and the mean years of education was 10.8 (SD=2.6). The mean cognitive scores at baseline and visit 6 were 70.6 (SD=6.0) and 65.4 (SD=9.6), respectively. BMI measures were collected starting from visit 1, and the mean BMI at visits 1 and 6 were 29.1 (SD=5.1) and 26.7 (SD=5.3), respectively. The histogram plots of cognitive function and BMI measures at baseline were also explored. Although CSID scores are slightly skewed toward lower scores, we assumed normal distributions for both CSID scores and BMI measures. We investigated the robustness of our proposed methods to non-normal distributions in simulation studies. Figure 2.1 shows the cognitive and BMI trajectories from 5 randomly selected IIDS participants. It can be seen from Figure 1 that, in general, both cognitive score and BMI decrease with age. We noted that these 238 participants used in our analysis are survivors with relatively long follow-up information and they expected to be healthier than others in the cohort who did not provide five measurements. In Section 2.8, we provide further discussion on the impact of missing data due to death and its potential impact on our analysis results.

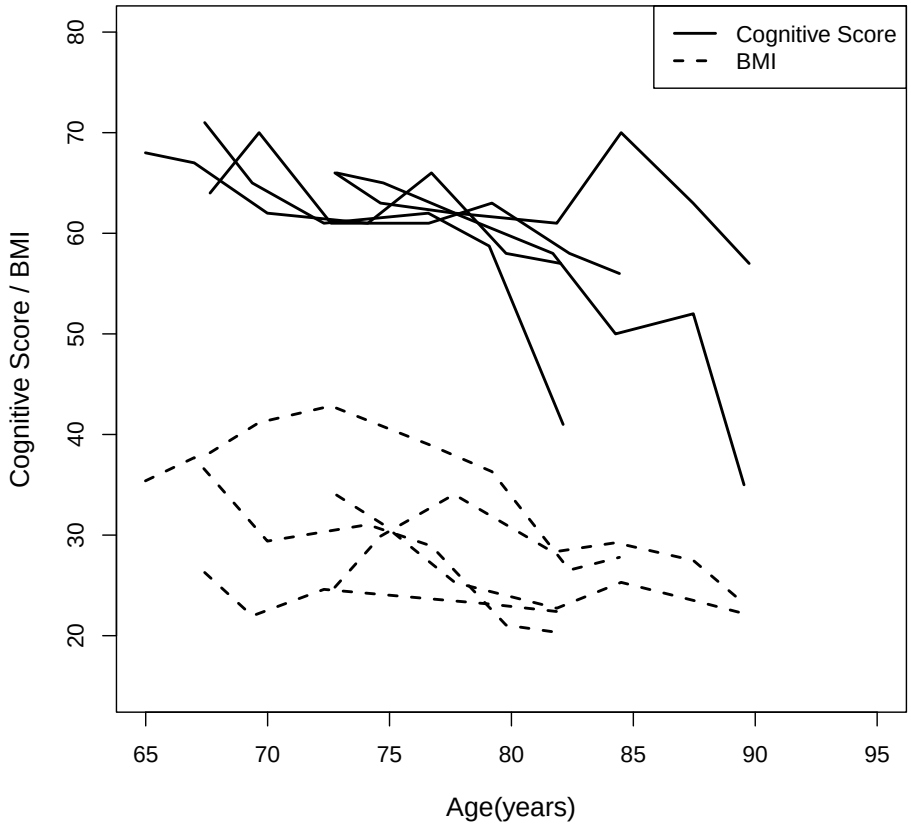

Figure 2.1: Observed longitudinal cognitive scores and BMI measures over time for five randomly selected participants from IIDS.

2.4 Statistical Methods

In this section, we define notations and introduce the three different random change point models for longitudinal outcomes. For each longitudinal outcome, we consider the random change point model with one change point that can be further extended to multiple change points. Let t_{ij} be the time of the j-th longitudinal measurement for the i-th subject, $i = 1, 2, ..., n,\ j = 1, 2, ..., m_i$; y_{1ij} and y_{2ij} are the bivariate longitudinal outcomes for the i-th subject at time t_{ij}.

2.4.1 Broken-Stick Model

For the i-th subject at time t_{ij},

$$y_{1ij} = \alpha_{1i} + \alpha_{2i}(t_{ij} - \alpha_{4i})I_{(-\infty,\alpha_{4i})}(t_{ij}) + \alpha_{3i}(t_{ij} - \alpha_{4i})I_{[\alpha_{4i},\infty)}(t_{ij}) + \epsilon_{1ij}, \quad (2.1)$$

$$y_{2ij} = \alpha_{5i} + \alpha_{6i}(t_{ij} - \alpha_{8i})I_{(-\infty,\alpha_{8i})}(t_{ij}) + \alpha_{7i}(t_{ij} - \alpha_{8i})I_{[\alpha_{8i},\infty)}(t_{ij}) + \epsilon_{2ij}, \quad (2.2)$$

where α_{4i} and α_{8i} denote the change points for y_{1ij} and y_{2ij}, respectively. α_{1i} and α_{5i} represent the intercepts in the two models and can be interpreted as the mean values of longitudinal outcomes at change points α_{4i} and α_{8i}, respectively. α_{2i} and α_{6i} denote the slopes before the change points, and α_{3i} and α_{7i} denote the slopes after the change points. ϵ_{1ij} and ϵ_{2ij} denote the residual errors of the longitudinal measurements, which are independently distributed as $\epsilon_{1ij} \sim_{iid} \mathrm{N}(0, \sigma^2_{e_1})$ and $\epsilon_{2ij} \sim_{iid} \mathrm{N}(0, \sigma^2_{e_2})$. $I_A(\cdot)$ is an indicator function with $I_A(x) = 1$ for $x \in A$ and $I_A(x) = 0$ for $x \notin A$.

In addition, we assume a multivariate distribution for the parameters in model 2.1 and 2.2,

$$\boldsymbol{\alpha}_i = (\alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \alpha_{4i}, \alpha_{5i}, \alpha_{6i}, \alpha_{7i}, \alpha_{8i})^T \sim \mathrm{MVN}(\alpha, \boldsymbol{\Sigma}_\alpha),$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8)^T$ is a 8×1 vector with each entry representing the population mean, and $\mathbf{\Sigma}_\alpha$ is the 8×8 variance-covariance matrix.

The broken-stick model can be implemented using a Bayesian framework and has simple parameter interpretation. However, it is not always appropriate because a sudden change in direction may not be a realistic assumption. Furthermore, the non-continuity at the change point can also cause numerical problems in parameter estimation using the frequentist method, such as the maximum likelihood method. Thus, there is a need to investigate other models not hampered by the disadvantages of the broken-stick model. Here, we use some IIDS data analysis results from Section 2.6 as an example to illustrate the three different models. In Figure 2.2, the black dots denote the cognitive function measures for a randomly selected individual and the black solid line illustrates the predicted broken-stick curve of this individual with a sudden transition happened at age of 78.13 years old.

2.4.2 Bacon-Watts Model

An alternative to the broken-stick model is the Bacon-Watts model (Bacon and Watts, 1971). For the i-th subject at time t_{ij},

$$y_{1ij} = \beta_{1i} + \beta_{2i}(t_{ij} - \beta_{4i}) + \beta_{3i}(t_{ij} - \beta_{4i})\text{trn}((t_{ij} - \beta_{4i})/\phi_1) + \epsilon_{1ij}, \tag{2.3}$$

$$y_{2ij} = \beta_{5i} + \beta_{6i}(t_{ij} - \beta_{8i}) + \beta_{7i}(t_{ij} - \beta_{8i})\text{trn}((t_{ij} - \beta_{8i})/\phi_2) + \epsilon_{2ij}, \tag{2.4}$$

where trn denotes the general transition function. Here, we choose to use the hyperbolic tangent function, tanh, a commonly used transition function; ϕ_1 and ϕ_2 are the transition parameters in the bivariate model and determine transition rates with larger values corresponding to slower transitions. In particular, if the transition parameter is close to

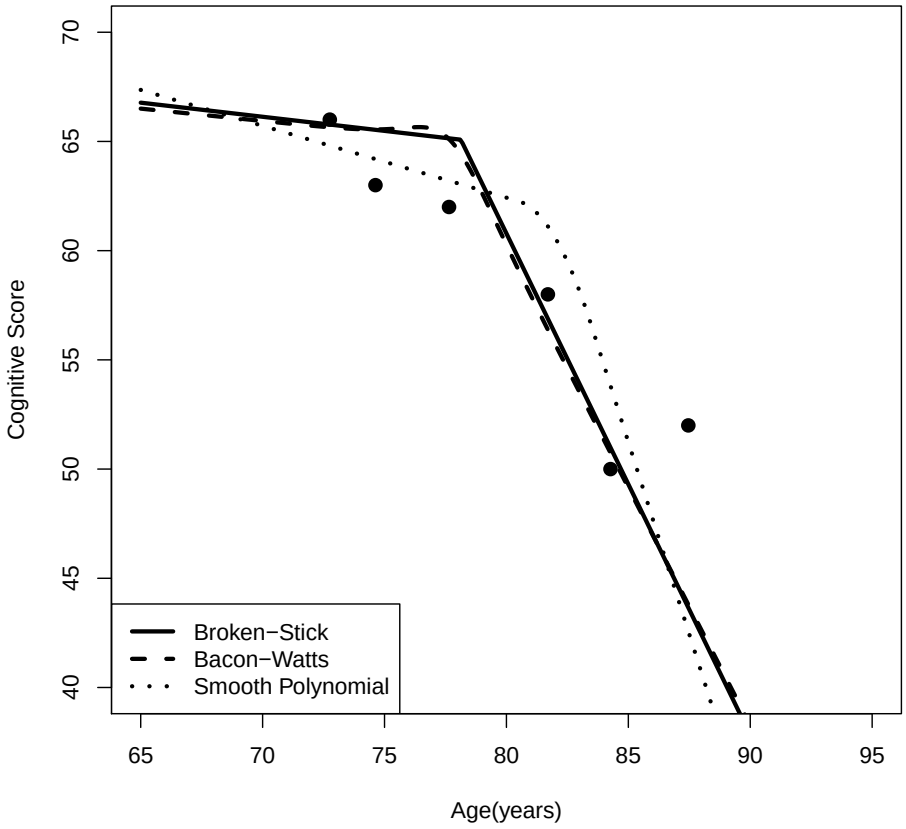

Figure 2.2: Predicted curves of the three types of change point model for the cognitive scores of an individual from IIDS.

zero, the Bacon-Watts model will work similarly to the broken-stick model. Parameters β_{1i} and β_{5i} denote the intercepts in each model, which have the same interpretation as in the random broken-stick models. Parameters β_{4i} and β_{8i} are change points in the bivariate model. However, the two slopes (β_{2i} and β_{3i}) in model 2.3 and the two slopes (β_{6i} and β_{7i}) in model 2.4 no longer have the same interpretation as in the random broken-stick model due to the formulation of the Bacon-Watts model. Again, we assume a multivariate normal distribution for all parameters in the bivariate model,

$$\beta_i = (\beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i}, \beta_{5i}, \beta_{6i}, \beta_{7i}, \beta_{8i})^T \sim \text{MVN}(\beta, \mathbf{\Sigma}_\beta),$$

where

$$\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8)^T,$$

is the vector of means, and $\mathbf{\Sigma}_\beta$ is the 8×8 variance-covariance matrix corresponding to the parameter vector.

Compared to the broken-stick model, the Bacon-Watts model enjoys continuity over the entire parameter space. However, its applicability may be limited because its slope parameters are difficult to interpret with respect to practice. Continuing the previous example in Section 2.4.1, the black dash line in Figure 2.2 shows the predicted Bacon-Watts curve with a smooth transition at the age of 77.68 years and a transition parameter of 1.60 for the selected subject.

2.4.3 Smooth Polynomial Model

Another alternative to the broken-stick model is the smooth polynomial model in which the continuity in the regions around the change points is achieved by using a polynomial function(van den Hout et al., 2010). The bivariate random smooth polynomial model for

the i-th subject at time t_{ij} is given by

$$y_{1ij} = (\eta_{1i} + \eta_{2i}t_{ij})I_{(-\infty,\eta_{4i})}(t_{ij}) + g_1(t_{ij}|\eta_{1i}, \eta_{2i}, \eta_{3i}, \varepsilon_1)I_{[\eta_{4i},\eta_{4i}+\varepsilon_1)}(t_{ij})$$
$$+(\lambda_{1i} + \eta_{3i}t_{ij})I_{[\eta_{4i}+\varepsilon_1,\infty)}(t_{ij}) + \epsilon_{1ij} \tag{2.5}$$

and

$$y_{2ij} = (\eta_{5i} + \eta_{6i}t_{ij})I_{(-\infty,\eta_{8i})}(t_{ij}) + g_2(t_{ij}|\eta_{5i}, \eta_{6i}, \eta_{7i}, \varepsilon_2)I_{[\eta_{8i},\eta_{8i}+\varepsilon_2)}(t_{ij})$$
$$+(\lambda_{2i} + \eta_{7i}t_{ij})I_{[\eta_{8i}+\varepsilon_2,\infty)}(t_{ij}) + \epsilon_{2ij} \tag{2.6}$$

where ε_1 and ε_2 denote the intervals around the change points that connect the two linear parts in each model and act as transition parameters as in the Bacon-Watts model but with a slightly different interpretation. As the transition parameter tends to zero, the interval around the change point tends to zero and the smooth polynomial model becomes the broken-stick model. Note that the parameters in the smooth polynomial models have different interpretations from the previous two models. The change points in the smooth polynomial models are defined as $\eta_{4i}+1/2\varepsilon_1$ and $\eta_{8i}+1/2\varepsilon_2$, respectively. η_{1i} and η_{5i} are the mean values of longitudinal measurements at η_{4i} and η_{8i} for the ith subject, respectively. Parameters η_{2i} and η_{3i} specify the slopes for the two linear parts before and after the smooth interval, respectively, for y_{1ij}, and η_{6i} and η_{7i} are defined similarly for y_{2ij}.

In model 2.5, λ_{1i} is derived by assuming the equality of the two linear parts at change point $\eta_{4i} + 1/2\varepsilon_1$; eventually it could be represented by a function of $(\eta_{1i}, \eta_{2i}, \eta_{3i}, \varepsilon_1)$. λ_{2i} in model 2.6 is derived by following the same argument. Hence,

$$\lambda_{1i} = \eta_{1i} + \eta_{2i}(\eta_{4i} + 1/2\varepsilon_1) - \eta_{3i}(\eta_{4i} + 1/2\varepsilon_1),$$

$$\lambda_{2i} = \eta_{5i} + \eta_{6i}(\eta_{8i} + 1/2\varepsilon_2) - \eta_{7i}(\eta_{8i} + 1/2\varepsilon_2).$$

g_1 and g_2 are two pre-specified polynomial functions that connect the two linear parts in each model. As in van den Hout et al. (2010), the smoothness of transition is achieved by imposing special constraints on g_1 so that the polynomial function will connect with the values of the linear function:

$$g_1(\eta_{4i}) = \eta_{1i} + \eta_{2i}\eta_{4i},\ g_1(\eta_{4i} + \varepsilon_1) = \eta_{1i} + \eta_{2i}(\eta_{4i} + \varepsilon_1),$$

$$(\frac{\partial}{t_{ij}}g_1)(\eta_{4i}) = \eta_{1i},\ (\frac{\partial}{t_{ij}}g_1)(\eta_{4i} + \varepsilon_1) = \eta_{2i}.$$

By defining g_1 as a cubic polynomial, $g_1(x) = a_3x^3 + a_2x^2 + a_1x + a_0$, and solving the above linear system of four linear ordinal differential equations, g_1 is a quadratic polynomial with the following coefficients (the coefficient of x^3 is zero):

$$a_2 = \frac{\eta_{3i} - \eta_{2i}}{2\varepsilon_1},\ a_1 = \eta_{2i} - \frac{\eta_{3i} - \eta_{2i}}{\varepsilon_1}\eta_{4i},\ a_0 = \eta_{1i} + \frac{\eta_{3i} - \eta_{2i}}{2\varepsilon_1}\eta_{3i}^2.$$

The form of g_2 can be specified similarly as g_1.

We again assume a multivariate normal distribution for all parameters in model 2.5 and 2.6:

$$\eta_i = (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{4i}, \eta_{5i}, \eta_{6i}, \eta_{7i}, \eta_{8i})^T \sim \text{MVN}(\eta, \mathbf{\Sigma}_\eta),$$

where

$$\eta = (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7, \eta_8)^T$$

represents the mean vector, and $\mathbf{\Sigma}_\eta$ is the 8×8 variance-covariance matrix corresponding to the parameter vector.

The smooth polynomial model not only maintains the advantages of the previous two models but also overcomes drawbacks of the previous two models. Thus the smooth polynomial model is superior in interpretable parameters and continuity at the change point. Again, in Figure 2.2, assuming a fixed interval of 3 years around the change point, the predicted smooth polynomial curve is illustrated (black dot line) for the selected individual. It is observed that the smooth curve started at 80.51 years old and the change point was at 82.01 years old, calculated by adding half of the interval (1.5 years) to 80.51.

2.4.4 Estimation Method

The maximum likelihood method is commonly used for parameter estimation in mixed-effects models. However, its use in models with multiple random effects can be challenging due to the need for multi-fold integrations. The Gaussian quadrature method, a numerical technique for approximating the multi-fold integration in mixed-effects models, can become computationally intractable when the number of random effects are large. In contrast, the Bayesian method using MCMC sampling avoids the direct multi-fold integration by taking repeated samplings from conditional posterior distribution for each parameter in the model, thus providing numerical solutions to a complex modeling situation.

The Bayesian method has been considered by Hall et al. (2003), Dominicus et al. (2008), van den Hout et al. (2010), and Hall et al. (2001) for parameter estimation from univariate random change point models. WinBUGS (Lunn et al., 2000) is a powerful and flexible statistical software for Bayesian inference using the Gibbs sampling technique. *BRugs* (Ligges et al., 2009) is a package in **R** (R Development Core Team, 2007) that also uses the Gibbs sampling method for Bayesian inference. *BRugs* performs similarly as WinBUGS with an additional advantage of combining data manipulation with the Bayesian model's fitting process including model specification and the choice of priors. We chose to implement our

methods using *BRugs* mostly because it can handle the simulations. For application to data analysis, we expect both WinBUGS and *BRugs* will be adequate for implementing the bivariate change point model.

The quality of fit is based on two criteria, the deviance information criterion (DIC) (Spiegelhalter et al., 2002) and the conditional predictive ordinate (CPO) (Gelfand et al., 1992). The trace plot of MCMC iterations is also monitored for purpose of convergence checking. The DIC has been widely used for Bayesian model comparison. Dominicus et al. (2008) used DIC to compare models with different structures as well as models differing in prior distributions. The DIC consists of two parts: DIC $= \bar{D} + p_D$, where $\bar{D}$ is the posterior expectation of deviance, and p_D is the effective number of parameters measuring the complexity of model (defined as the posterior mean of the deviance minus the deviance of the posterior means). Similar to Akaike information criterion (AIC) (Akaike, 1987), a smaller DIC corresponds to a better fit. Another frequently used model-selection criteria in Bayesian inference is the CPO, a cross-validated predictive approach calculating the predictive distributions conditioned on the observed data by leaving out one observation each time. Chen et al. (2000) showed that there existed a Monte Carlo approximation of the CPO. The models are compared using the log pseudo-marginal likelihood (LPML), which is defined as LPML $= \sum_{i=1}^{n} \log(\widehat{\text{CPO}}_i)$, where n is the total number of observations and $\widehat{\text{CPO}}_i$ is the Monte Carlo approximation of CPO. Contrary to the DIC, the model with larger LPML indicates a better fit.

2.5 Simulation Study

We used Monte Carlo (MC) simulations to assess the performance of the Bayesian approach for parameter estimation in the proposed bivariate random smooth polynomial models because the smooth polynomial model is more realistic in practice and more comprehensive

than the other two models. We simulated data from a bivariate random smooth polynomial model using the estimated parameters from fitting this model to the real data (IIDS). Specifically, each simulated MC data set consists of bivariate longitudinal data from 238 subjects with 7 non-missing bivariate repeated measurements per subject (equally spaced with 3 years between the two adjacent visits). Baseline ages from IIDS subtracted by 65 years were used as ages at the first visit for each subject.

We present simulation results for 12 scenarios by varying the correlation between the two change points, variances of change points and measurement errors (Table 2.1):

Table 2.1: Considered 12 simulation scenarios differing in correlation between two change points ($r_{\eta_4\eta_8}$), variance of each change point ($\sigma^2_{\eta_4}, \sigma^2_{\eta_8}$) and variance of each measurement error ($\sigma^2_{\epsilon_1}, \sigma^2_{\epsilon_2}$).

Scenario	$r_{\eta_4\eta_8}$	$\sigma^2_{\eta_4}, \sigma^2_{\eta_8}$	$\sigma^2_{\epsilon_1}, \sigma^2_{\epsilon_2}$
1	0.2	64, 16	20, 5
2	0.2	64, 16	5, 1
3	0.2	16, 4	20, 5
4	0.2	16, 4	5, 1
5	0.4	64, 16	20, 5
6	0.4	64, 16	5, 1
7	0.4	16, 4	20, 5
8	0.4	16, 4	5, 1
9	0.6	64, 16	20, 5
10	0.6	64, 16	5, 1
11	0.6	16, 4	20, 5
12	0.6	16, 4	5, 1

The other parameters were chosen to be close to the estimated parameters from IIDS data: $\eta_1 = 70$, $\sigma^2_{\eta_1} = 15$, $\eta_2 = -0.2$, $\sigma^2_{\eta_2} = 0.2$, $\eta_3 = -3$, $\sigma^2_{\eta_3} = 2$, $\eta_4 = 15$, $\eta_5 = 28$, $\sigma^2_{\eta_5} = 16$, $\eta_6 = 0.2$, $\sigma^2_{\eta_6} = 0.2$, $\eta_7 = -0.4$, $\sigma^2_{\eta_7} = 0.2$, $\eta_8 = 10$, $r_{\eta_2\eta_3} = 0.2$, $r_{\eta_6\eta_7} = -0.5$, $r_{\eta_4\eta_8} = 0.4$, $\varepsilon_1 = 3$, and $\varepsilon_2 = 3$. Here, $r_{\eta_2\eta_3}$ denoted the correlation between η_2 and η_3, $r_{\eta_6\eta_7}$ and $r_{\eta_4\eta_8}$ were defined similarly. Thus, in the 8 by 8 variance-covariance matrix $\mathbf{\Sigma}_\eta$,

only $\sigma_{\eta_2\eta_3}$, $\sigma_{\eta_6\eta_7}$ and $\sigma_{\eta_4\eta_8}$ were nonzero, and all the other off-diagonal elements were set to be zeros.

2.5.1 Estimation Using Bivariate Random Smooth Polynomial Models

In the Bayesian model fitting of the bivariate random smooth polynomial model, prior distributions of parameters for scenario 1 were chosen as the following:

$$\eta_1 \sim \mathrm{N}(65, 0.01), \eta_5 \sim \mathrm{N}(25, 0.01),$$

$$\sigma^2_{\eta_1} \sim \mathrm{invGamma}(0.001, 0.001),$$

$$\begin{pmatrix} \eta_2 \\ \eta_3 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix} \right),$$

$$\mathbf{\Sigma}_{\eta_2\eta_3} \sim \mathrm{invWishart}\left(\begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}, \ 2 \right),$$

$$\begin{pmatrix} \eta_4 \\ \eta_8 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 15 \\ 10 \end{pmatrix}, \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix} \right),$$

$$\mathbf{\Sigma}_{\eta_4\eta_8} \sim \mathrm{invWishart}\left(\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}, \ 2 \right),$$

$(\eta_6, \eta_7)^T$ and $\mathbf{\Sigma}_{\eta_6\eta_7}$ have the same prior distribution as $(\eta_2, \eta_3)^T$ and $\mathbf{\Sigma}_{\eta_2\eta_3}$, respectively; $\sigma^2_{\eta_5}$, $\sigma^2_{\epsilon_1}$ and $\sigma^2_{\epsilon_2}$ also have the same priors as $\sigma^2_{\eta_1}$. In Bayesian analysis, in particular, conjugate prior is a natural and popular choice because of its flexibility and mathematical convenience. invGamma(α, β) was chosen as it is commonly used as the conjugate prior

to the variance of univariate normal distribution, where α is the shape parameter and β is the scale parameter. On the other hand, invWishart$(\mathbf{\Sigma}, k)$ was a conjugate prior to the variance-covariance matrix of a multivariate normal distribution, where $\mathbf{\Sigma}$ is a positive definite inverse scale matrix and the positive integer k denotes the degree of freedom. Priors for scenario 2 were chosen the same as in scenario 1 except the variance-covariance of the two change points:

$$\mathbf{\Sigma}_{\eta_4\eta_8} \sim \text{invWishart}\left(\begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}, \quad 2\right).$$

The two transition parameters ε_1 and ε_2 were first treated as fixed (equal to the true values) in the model fitting for the two scenarios. For each scenario, 500 MC samples were generated and fitted by the bivariate random smooth polynomial model. For each MC sample, $20,000$ additional iterations were considered following 2000 burn-in iterations. The simulation results are presented in Table 2.2, 2.3, 2.4, 2.5, 2.6 and 2.7. For each scenario, we reported mean, mean squared error, mean standard error, empirical standard error, and coverage probabilities of 95% posterior intervals. The simulation results showed that the Bayesian method generally performed well for fitting the bivariate smooth random polynomial model: estimated parameters had low bias and coverage probability rates of 95% posterior credible intervals were around the nominal level. It is also observed that model-fitting is influenced by the variances of change points and variance of measurement errors. Specifically, smaller variances of change points or variance of measurement errors led to parameter estimates with smaller bias, as well as smaller MSEs. We also conducted a simulation study treating the two transition parameters as unknown parameters and setting uniform prior distributions for them. The simulation results are presented in Table 2.8 and 2.9. We found few differences in parameter estimations between the two situations.

2.5.2 Estimation Using Broken-Stick and Bacon-Watts Models

We have been focusing on investigating the performance of the bivariate random smooth polynomial model via simulation studies. However, the random smooth polynomial model is much more complex in model structure than the other two models, and consequently more computationally expensive in practice; thus there is a need to study the performance of the other two simplified bivariate models under the assumption that the true model is the bivariate random smooth polynomial model.

Prior distributions for the bivariate random broken-stick model and the bivariate random Bacon-Watts model were chosen similarly to that in the bivariate random smooth polynomial model. The two transition parameters in the bivariate random Bacon-Watts model were treated as unknown parameters with uniform prior distributions,

$$\phi_1 \sim \text{Unif}(0.1, 5),$$

$$\phi_2 \sim \text{Unif}(0.1, 5).$$

Table 2.10, 2.11, 2.12, 2.13, 2.14 and 2.15 summarized the simulation results of the three different bivariate models for the 12 scenarios. Since most model parameters were not directly comparable due to different model parameterizations, only the following parameters were compared among the three bivariate models: change points, variances of change points, and correlations between change points. Under the assumption that the true model is a bivariate random smooth polynomial model, simulation results confirmed that the bivariate random smooth polynomial model had the best performance among the three modeling frameworks with smaller bias, smaller MSEs, and better posterior interval coverage. In contrast, the bivariate random broken-stick model and the bivariate random Bacon-Watts

model showed larger bias, larger MSEs and worse posterior interval coverage in parameter estimations than those obtained under the bivariate random smooth polynomial model.

The bivariate random broken-stick model and the bivariate random Bacon-Watts model had similar parameter estimation results. When the variances of random change points were larger, the change points were underestimated by approximately two years; estimates of variances of change points and correlation between change points also deviated from the true values. However, when the variances of measurement error and variances of change points were small, the bivariate random broken-stick model and the bivariate random Bacon-Watts model had much improved performance, nearly as well as the bivariate random smooth polynomial model.

2.5.3 Sensitivity Analysis

Estimation of change point of random change point model is usually sensitive to the distributional assumption of the data. To study the sensitivity and robustness of the proposed methods, we replaced the normal distribution in generating the simulated data for random effects and error density by lognormal distributions. Again, for each scenario, 500 Monte Carlo samples, each with 238 subjects and 7 non-missing bivariate repeated measurements per subject, were generated from the bivariate random smooth polynomial model with correlated slopes in each univariate model. The true parameters were selected to be the same as those in the previous simulation study, and then transformed to the mean and standard deviations of the lognormal distribution to ensure the generated data have the similar range as when using normal distributions. The generated MC samples were then fitted by the 3 bivariate change point models assuming normal distributions of the random effect and error terms. The prior distributions for parameters for each model were chosen in similar fashion as in the previous section. For each MC sample, $20,000$ additional iterations were consid-

ered following 2000 burn-in iterations. Simulation results, including the estimates of change points, variances of change points, and the correlation estimates between change points are presented in Table 2.16. The results show that for both of the small and large scenario the bivariate random smooth polynomial model has the best performance in smaller MSE and better 95% PI coverage. It is also observed that under the assumption of the lognormal distribution and the smooth polynomial model, the bivariate random broken-stick model and the bivariate Bacon-Watts model are sensitive in estimating change points. In addition, the variance of change points and measurement errors have some impact on the model-fitting results. Specifically, the change point estimates from the small variance scenario are less sensitive than those from the large variance scenario.

Table 2.2: Simulation results of bivariate random smooth polynomial model under scenarios 1 and 2.

	Scenario 1						Scenario 2					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.97	0.2130	0.46	0.46	95.4%	70.0	69.98	0.1150	0.33	0.34	94.8%
$\sigma^2_{\eta_1}$	15.0	15.28	8.3417	2.88	2.88	95.6%	15.0	15.14	3.9860	2.00	1.99	93.8%
η_2	-0.2	-0.20	0.0021	0.05	0.05	96%	-0.2	-0.20	0.0010	0.03	0.03	95.8%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0007	0.03	0.03	94.2%	0.1	0.10	0.0013	0.02	0.04	93.4%
η_3	-3.0	-3.02	0.0438	0.20	0.21	94%	-3.0	-3.01	0.0234	0.15	0.15	94.8%
$\sigma^2_{\eta_3}$	2.0	1.92	0.1454	0.38	0.37	95.6%	2.0	1.98	0.0876	0.31	0.30	95.2%
η_4	15.0	15.07	1.0849	0.99	1.04	93.4%	15.0	15.04	0.5748	0.73	0.76	95.2%
$\sigma^2_{\eta_4}$	64.0	66.11	124.1853	11.21	10.95	94.6%	64.0	65.53	86.8044	9.27	9.20	95.2%
η_5	28.0	28.01	0.1408	0.38	0.38	95.8%	28.0	28.00	0.0974	0.31	0.31	94.6%
$\sigma^2_{\eta_5}$	16.0	16.42	5.8691	2.32	2.39	92.4%	16.0	16.23	3.6574	1.89	1.90	93.8%
η_6	0.2	0.21	0.0033	0.05	0.06	93.2%	0.2	0.20	0.0016	0.04	0.04	94.6%
$\sigma^2_{\eta_6}$	0.2	0.19	0.0012	0.03	0.03	94.8%	0.2	0.20	0.0006	0.03	0.03	94.8%
η_7	-0.4	-0.41	0.0018	0.04	0.04	94.8%	-0.4	-0.40	0.0012	0.03	0.03	93.8%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0008	0.03	0.03	95.2%	0.2	0.20	0.0004	0.02	0.02	96.2%
η_8	10.0	9.87	0.6644	0.85	0.81	95.8%	10.0	9.95	0.2454	0.52	0.49	94.8%
$\sigma^2_{\eta_8}$	16.0	20.86	45.9940	5.41	4.73	88.6%	16.0	18.23	16.3284	3.41	3.38	92%
$r_{\eta_4\eta_8}$	0.2	0.17	0.0221	0.15	0.15	96%	0.2	0.18	0.0137	0.12	0.12	96.2%
$\sigma^2_{\epsilon_1}$	20.0	20.11	0.8320	0.88	0.91	94.8%	4.0	4.02	0.0359	0.18	0.19	94.8%
$\sigma^2_{\epsilon_2}$	5.0	5.04	0.0524	0.23	0.23	94.4%	1.0	1.01	0.0025	0.05	0.05	93.6%
$r_{\eta_2\eta_3}$	0.2	0.17	0.0702	0.24	0.26	91%	0.2	0.19	0.0247	0.15	0.16	93.4%
$r_{\eta_6\eta_7}$	-0.5	-0.56	0.0169	0.12	0.12	91%	-0.5	-0.52	0.0076	0.08	0.08	92.8%

Table 2.3: Simulation results of bivariate random smooth polynomial model under scenarios 3 and 4.

	Scenario 3						Scenario 4					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.95	0.2269	0.46	0.47	94.8%	70.0	69.97	0.1145	0.32	0.34	95%
$\sigma^2_{\eta_1}$	15.0	15.30	8.0407	2.77	2.82	94.8%	15.0	15.09	3.7977	1.93	1.95	94.2%
η_2	-0.2	-0.20	0.0020	0.05	0.04	95.4%	-0.2	-0.20	0.0008	0.03	0.03	95.8%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0006	0.02	0.02	93.6%	0.1	0.10	0.0002	0.01	0.02	94.6%
η_3	-3.0	-3.01	0.0324	0.17	0.18	94%	-3.0	-3.01	0.0197	0.13	0.14	93.6%
$\sigma^2_{\eta_3}$	2.0	1.96	0.1130	0.33	0.33	94%	2.0	1.99	0.0680	0.27	0.26	95.4%
η_4	15.0	15.01	0.2765	0.50	0.53	94.4%	15.0	15.02	0.1327	0.36	0.36	94.2%
$\sigma^2_{\eta_4}$	16.0	16.33	10.3572	3.04	3.20	92.2%	16.0	16.26	5.5326	2.30	2.34	93.4%
η_5	28.0	28.00	0.1335	0.37	0.37	96%	28.0	27.99	0.0940	0.31	0.31	95.6%
$\sigma^2_{\eta_5}$	16.0	16.41	5.6724	2.30	2.35	92.8%	16.0	16.23	3.6652	1.88	1.90	94.8%
η_6	0.2	0.20	0.0022	0.05	0.05	94%	0.2	0.20	0.0013	0.04	0.04	94.8%
$\sigma^2_{\eta_6}$	0.2	0.20	0.0010	0.03	0.03	93.8%	0.2	0.20	0.0006	0.02	0.02	94%
η_7	-0.4	-0.40	0.0015	0.04	0.04	94.2%	-0.4	-0.40	0.0010	0.03	0.03	94.8%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0007	0.03	0.03	95.8%	0.2	0.20	0.0004	0.02	0.02	97.6%
η_8	10.0	10.01	0.2285	0.48	0.48	94%	10.0	10.02	0.0702	0.27	0.26	95.6%
$\sigma^2_{\eta_8}$	4.0	4.63	3.1355	1.73	1.66	96.2%	4.0	4.19	0.9641	0.98	0.96	95.2%
$r_{\eta_4\eta_8}$	0.2	0.18	0.0465	0.22	0.22	94.4%	0.2	0.19	0.0194	0.14	0.14	94%
$\sigma^2_{\epsilon_1}$	20.0	20.11	0.8424	0.89	0.91	95.4%	4.0	4.01	0.0367	0.19	0.19	95%
$\sigma^2_{\epsilon_2}$	5.0	5.03	0.0502	0.22	0.22	95.2%	1.0	1.01	0.0022	0.05	0.05	94.6%
$r_{\eta_2\eta_3}$	0.2	0.22	0.0362	0.18	0.19	93.2%	0.2	0.20	0.0150	0.12	0.12	92.8%
$r_{\eta_6\eta_7}$	-0.5	-0.50	0.0087	0.09	0.09	94.4%	-0.5	-0.50	0.0047	0.07	0.07	94%

Table 2.4: Simulation results of bivariate random smooth polynomial model under scenarios 5 and 6.

	Scenario 5						Scenario 6					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.97	0.2129	0.46	0.46	95.2%	70.0	69.98	0.1143	0.33	0.34	95%
$\sigma^2_{\eta_1}$	15.0	15.33	8.5522	2.88	2.91	95.6%	15.0	15.15	4.0122	1.99	2.00	93.6%
η_2	-0.2	-0.20	0.0021	0.05	0.05	96%	-0.2	-0.20	0.0010	0.03	0.03	95.4%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0009	0.03	0.03	94.8%	0.1	0.10	0.0013	0.02	0.04	92.4%
η_3	-3.0	-3.02	0.0424	0.20	0.20	95%	-3.0	-3.01	0.0234	0.15	0.15	94.8%
$\sigma^2_{\eta_3}$	2.0	1.92	0.1477	0.38	0.38	95%	2.0	1.97	0.0874	0.30	0.29	95.6%
η_4	15.0	15.08	1.0395	0.98	1.02	94.2%	15.0	15.06	0.5823	0.72	0.76	94.8%
$\sigma^2_{\eta_4}$	64.0	66.14	121.5335	11.16	10.83	95.4%	64.0	65.57	85.0290	9.20	9.10	95.6%
η_5	28.0	28.01	0.1404	0.38	0.37	96%	28.0	28.00	0.0965	0.31	0.31	94.6%
$\sigma^2_{\eta_5}$	16.0	16.44	5.8107	2.31	2.37	92.8%	16.0	16.21	3.6853	1.89	1.91	94.4%
η_6	0.2	0.22	0.0033	0.05	0.06	92.6%	0.2	0.21	0.0016	0.04	0.04	95%
$\sigma^2_{\eta_6}$	0.2	0.19	0.0012	0.03	0.03	94.6%	0.2	0.20	0.0006	0.03	0.03	95.6%
η_7	-0.4	-0.41	0.0018	0.04	0.04	94.2%	-0.4	-0.40	0.0012	0.03	0.03	93.4%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0008	0.03	0.03	94.4%	0.2	0.20	0.0005	0.02	0.02	96.4%
η_8	10.0	9.87	0.6481	0.83	0.79	95%	10.0	9.95	0.2252	0.50	0.47	96.2%
$\sigma^2_{\eta_8}$	16.0	21.03	47.5446	5.26	4.72	87.2%	16.0	18.28	15.9585	3.36	3.28	91.2%
$r_{\eta_4\eta_8}$	0.4	0.33	0.0223	0.14	0.13	95%	0.4	0.36	0.0125	0.11	0.10	95.6%
$\sigma^2_{\epsilon_1}$	20.0	20.11	0.8386	0.88	0.91	94.8%	4.0	4.02	0.0359	0.18	0.19	94.4%
$\sigma^2_{\epsilon_2}$	5.0	5.04	0.0523	0.23	0.23	95%	1.0	1.01	0.0024	0.05	0.05	93.2%
$r_{\eta_2\eta_3}$	0.2	0.18	0.0684	0.24	0.26	93%	0.2	0.19	0.0251	0.15	0.16	92.6%
$r_{\eta_6\eta_7}$	-0.5	-0.56	0.0174	0.12	0.12	90.4%	-0.5	-0.52	0.0075	0.08	0.08	92.8%

Table 2.5: Simulation results of bivariate random smooth polynomial model under scenarios 7 and 8.

	Scenario 7						Scenario 8					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.96	0.2280	0.46	0.48	94.6%	70.0	69.97	0.1145	0.32	0.34	95.2%
$\sigma^2_{\eta_1}$	15.0	15.30	7.9040	2.77	2.80	95.8%	15.0	15.09	3.7890	1.92	1.95	94%
η_2	-0.2	-0.20	0.0020	0.05	0.05	95.2%	-0.2	-0.20	0.0008	0.03	0.03	96.2%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0005	0.02	0.02	94%	0.1	0.10	0.0002	0.01	0.02	94.8%
η_3	-3.0	-3.01	0.0319	0.17	0.18	94.8%	-3.0	-3.01	0.0197	0.13	0.14	94.6%
$\sigma^2_{\eta_3}$	2.0	1.96	0.1114	0.33	0.33	94.2%	2.0	1.99	0.0671	0.27	0.26	95.4%
η_4	15.0	15.01	0.2760	0.50	0.53	93.4%	15.0	15.02	0.1321	0.35	0.36	94.6%
$\sigma^2_{\eta_4}$	16.0	16.36	10.3769	3.02	3.20	92%	16.0	16.24	5.4729	2.28	2.33	93.8%
η_5	28.0	28.00	0.1333	0.37	0.37	95.6%	28.0	28.00	0.0953	0.31	0.31	95%
$\sigma^2_{\eta_5}$	16.0	16.42	5.6975	2.30	2.35	93.2%	16.0	16.23	3.6598	1.88	1.90	94.2%
η_6	0.2	0.20	0.0022	0.05	0.05	93.8%	0.2	0.20	0.0013	0.04	0.04	93.4%
$\sigma^2_{\eta_6}$	0.2	0.20	0.0010	0.03	0.03	93.8%	0.2	0.20	0.0006	0.02	0.02	94.2%
η_7	-0.4	-0.40	0.0015	0.04	0.04	93.6%	-0.4	-0.40	0.0010	0.03	0.03	95%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0007	0.03	0.03	96.2%	0.2	0.20	0.0004	0.02	0.02	97.2%
η_8	10.0	10.01	0.2211	0.47	0.47	95%	10.0	10.02	0.0684	0.27	0.26	94.4%
$\sigma^2_{\eta_8}$	4.0	4.73	3.0068	1.70	1.57	96.4%	4.0	4.20	0.9320	0.97	0.94	95.6%
$r_{\eta_4\eta_8}$	0.4	0.35	0.0394	0.20	0.19	95.6%	0.4	0.39	0.0156	0.13	0.12	95.8%
$\sigma^2_{\epsilon_1}$	20.0	20.11	0.8416	0.89	0.91	95.4%	4.0	4.01	0.0364	0.19	0.19	95.6%
$\sigma^2_{\epsilon_2}$	5.0	5.03	0.0501	0.22	0.22	95.2%	1.0	1.01	0.0022	0.05	0.05	94.6%
$r_{\eta_2\eta_3}$	0.2	0.22	0.0363	0.18	0.19	93.8%	0.2	0.20	0.0150	0.11	0.12	92.4%
$r_{\eta_6\eta_7}$	-0.5	-0.51	0.0084	0.09	0.09	94.8%	-0.5	-0.50	0.0047	0.07	0.07	93.6%

Table 2.6: Simulation results of bivariate random smooth polynomial model under scenarios 9 and 10.

	Scenario 9						Scenario 10					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.98	0.2149	0.46	0.46	95.2%	70.0	69.98	0.1134	0.33	0.34	95.2%
$\sigma^2_{\eta_1}$	15.0	15.33	8.6280	2.88	2.92	95.4%	15.0	15.15	3.8999	1.99	1.97	94%
η_2	-0.2	-0.20	0.0021	0.05	0.05	97%	-0.2	-0.20	0.0010	0.03	0.03	95.4%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0007	0.03	0.03	94%	0.1	0.10	0.0011	0.02	0.03	92%
η_3	-3.0	-3.03	0.0428	0.20	0.20	94.6%	-3.0	-3.02	0.0227	0.15	0.15	94.8%
$\sigma^2_{\eta_3}$	2.0	1.91	0.1440	0.37	0.37	94.8%	2.0	1.96	0.0868	0.30	0.29	95%
η_4	15.0	15.12	1.0207	0.97	1.00	94%	15.0	15.09	0.5566	0.71	0.74	94.8%
$\sigma^2_{\eta_4}$	64.0	66.51	124.1506	11.09	10.87	95%	64.0	65.78	84.1441	9.11	9.01	95.6%
η_5	28.0	28.01	0.1405	0.38	0.37	95.2%	28.0	28.00	0.0951	0.31	0.31	94.4%
$\sigma^2_{\eta_5}$	16.0	16.44	5.7475	2.30	2.36	92.6%	16.0	16.21	3.6762	1.88	1.91	93.8%
η_6	0.2	0.22	0.0031	0.05	0.05	93.4%	0.2	0.21	0.0015	0.04	0.04	95.8%
$\sigma^2_{\eta_6}$	0.2	0.19	0.0011	0.03	0.03	95%	0.2	0.20	0.0006	0.03	0.02	95.6%
η_7	-0.4	-0.41	0.0017	0.04	0.04	94.2%	-0.4	-0.40	0.0011	0.03	0.03	93.6%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0008	0.03	0.03	95.2%	0.2	0.20	0.0004	0.02	0.02	95.6%
η_8	10.0	9.87	0.5499	0.79	0.73	95.6%	10.0	9.95	0.2007	0.47	0.45	96.6%
$\sigma^2_{\eta_8}$	16.0	21.30	48.3244	5.10	4.51	84.6%	16.0	18.38	15.3517	3.26	3.11	91.2%
$r_{\eta_4\eta_8}$	0.6	0.48	0.0259	0.12	0.11	88.8%	0.6	0.54	0.0106	0.09	0.08	91.4%
$\sigma^2_{\epsilon_1}$	20.0	20.11	0.8430	0.88	0.91	94.8%	4.0	4.02	0.0365	0.18	0.19	94.6%
$\sigma^2_{\epsilon_2}$	5.0	5.03	0.0513	0.22	0.22	95.4%	1.0	1.01	0.0023	0.05	0.05	93.8%
$r_{\eta_2\eta_3}$	0.2	0.18	0.0691	0.24	0.26	93.6%	0.2	0.19	0.0252	0.15	0.16	93%
$r_{\eta_6\eta_7}$	-0.5	-0.57	0.0169	0.11	0.11	90.2%	-0.5	-0.52	0.0072	0.08	0.08	93.2%

Table 2.7: Simulation results of bivariate random smooth polynomial model under scenarios 11 and 12.

	Scenario 11						Scenario 12					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.96	0.2274	0.45	0.48	94.6%	70.0	69.97	0.1158	0.32	0.34	94.8%
$\sigma^2_{\eta_1}$	15.0	15.31	7.9145	2.77	2.80	95%	15.0	15.09	3.7511	1.92	1.94	94.4%
η_2	-0.2	-0.20	0.0020	0.04	0.04	95.2%	-0.2	-0.20	0.0008	0.03	0.03	96%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0005	0.02	0.02	93.8%	0.1	0.10	0.0002	0.01	0.02	94.4%
η_3	-3.0	-3.01	0.0310	0.17	0.18	94.2%	-3.0	-3.01	0.0191	0.13	0.14	94%
$\sigma^2_{\eta_3}$	2.0	1.95	0.1102	0.33	0.33	93.6%	2.0	1.99	0.0666	0.27	0.26	94.8%
η_4	15.0	15.02	0.2641	0.49	0.51	94.2%	15.0	15.02	0.1291	0.35	0.36	94%
$\sigma^2_{\eta_4}$	16.0	16.40	10.1280	3.02	3.16	92.4%	16.0	16.25	5.3175	2.26	2.29	93.6%
η_5	28.0	28.00	0.1335	0.37	0.37	95.4%	28.0	28.00	0.0947	0.31	0.31	95.2%
$\sigma^2_{\eta_5}$	16.0	16.43	5.7281	2.30	2.36	93%	16.0	16.24	3.6439	1.88	1.90	94%
η_6	0.2	0.20	0.0022	0.05	0.05	94%	0.2	0.20	0.0013	0.03	0.04	94.4%
$\sigma^2_{\eta_6}$	0.2	0.20	0.0010	0.03	0.03	94%	0.2	0.20	0.0006	0.02	0.02	94.4%
η_7	-0.4	-0.40	0.0015	0.04	0.04	94.2%	-0.4	-0.40	0.0010	0.03	0.03	94.8%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0007	0.03	0.03	96%	0.2	0.20	0.0004	0.02	0.02	97.2%
η_8	10.0	10.01	0.2152	0.47	0.46	96%	10.0	10.02	0.0640	0.26	0.25	95.8%
$\sigma^2_{\eta_8}$	4.0	4.93	3.1687	1.70	1.52	95.6%	4.0	4.28	0.9083	0.94	0.91	94.8%
$r_{\eta_4\eta_8}$	0.6	0.50	0.0340	0.17	0.16	95.8%	0.6	0.57	0.0112	0.11	0.10	95.6%
$\sigma^2_{\epsilon_1}$	20.0	20.11	0.8373	0.89	0.91	95.2%	4.0	4.01	0.0369	0.19	0.19	94.8%
$\sigma^2_{\epsilon_2}$	5.0	5.02	0.0496	0.22	0.22	95.6%	1.0	1.00	0.0022	0.05	0.05	95.4%
$r_{\eta_2\eta_3}$	0.2	0.21	0.0361	0.18	0.19	94.2%	0.2	0.20	0.0149	0.11	0.12	92.6%
$r_{\eta_6\eta_7}$	-0.5	-0.51	0.0080	0.09	0.09	95.6%	-0.5	-0.50	0.0045	0.06	0.07	94%

Table 2.8: Simulation results of scenarios 5 and 6 for bivariate random smooth polynomial model with unknown ε_1 and ε_2.

	Scenario 5						Scenario 6					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.97	0.2125	0.46	0.46	95.6%	70.0	69.98	0.1143	0.33	0.34	94.8%
$\sigma^2_{\eta_1}$	15.0	15.29	8.3638	2.88	2.88	95.4%	15.0	15.23	4.8263	2.00	2.19	92.8%
η_2	-0.2	-0.20	0.0021	0.05	0.05	96.8%	-0.2	-0.20	0.0011	0.03	0.03	94%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0009	0.03	0.03	94%	0.1	0.10	0.0012	0.02	0.03	94%
η_3	-3.0	-3.03	0.0435	0.20	0.21	95%	-3.0	-3.01	0.0235	0.15	0.15	96.4%
$\sigma^2_{\eta_3}$	2.0	1.91	0.1460	0.38	0.37	95.6%	2.0	2.00	0.0889	0.31	0.30	96.4%
η_4	15.0	15.12	1.1597	1.18	1.07	97%	15.0	15.06	0.6775	0.84	0.82	96%
$\sigma^2_{\eta_4}$	64.0	66.47	124.5624	11.29	10.89	95.2%	64.0	65.95	84.1723	9.31	8.97	96%
η_5	28.0	28.01	0.1411	0.38	0.38	95.4%	28.0	28.01	0.0945	0.31	0.31	95.6%
$\sigma^2_{\eta_5}$	16.0	16.43	5.8169	2.31	2.38	92.8%	16.0	16.12	3.3271	1.88	1.82	95.2%
η_6	0.2	0.22	0.0033	0.05	0.06	93.6%	0.2	0.21	0.0015	0.04	0.04	96%
$\sigma^2_{\eta_6}$	0.2	0.19	0.0012	0.03	0.03	95%	0.2	0.20	0.0006	0.03	0.02	97.4%
η_7	-0.4	-0.41	0.0018	0.04	0.04	94.6%	-0.4	-0.40	0.0012	0.03	0.03	93.2%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0008	0.03	0.03	95%	0.2	0.20	0.0005	0.02	0.02	94.4%
η_8	10.0	10.00	0.6513	1.10	0.81	99.6%	10.0	10.00	0.3925	0.80	0.63	98.2%
$\sigma^2_{\eta_8}$	16.0	21.10	48.3310	5.28	4.73	88%	16.0	18.31	15.7782	3.36	3.24	90.8%
$r_{\eta_4\eta_8}$	0.4	0.33	0.0225	0.14	0.13	95%	0.4	0.36	0.0119	0.11	0.10	95.6%
$\sigma^2_{\epsilon_1}$	20.0	20.12	0.8392	0.88	0.91	94.8%	4.0	4.01	0.0353	0.19	0.19	94.4%
$\sigma^2_{\epsilon_2}$	5.0	5.03	0.0521	0.23	0.23	95%	1.0	1.00	0.0024	0.05	0.05	94.6%
ε_1	3.0	2.98	0.3192	1.26	0.57	100%	3.0	3.02	0.6513	0.85	0.81	92.6%
ε_2	3.0	2.68	0.2576	1.36	0.40	100%	3.0	2.86	0.6040	1.13	0.77	98%
$r_{\eta_2\eta_3}$	0.2	0.16	0.0714	0.25	0.27	93%	0.2	0.21	0.0275	0.15	0.17	92.4%
$r_{\eta_6\eta_7}$	-0.5	-0.56	0.0170	0.12	0.11	91.2%	-0.5	-0.52	0.0076	0.08	0.08	92.8%

Table 2.9: Simulation results of scenarios 7 and 8 for bivariate random smooth polynomial model with unknown ε_1 and ε_2.

	Scenario 7						Scenario 8					
	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage	True	Mean	MSE	Mean SD	Empirical SD	95% PI Coverage
η_1	70.0	69.94	0.2211	0.46	0.47	94.4%	70.0	69.96	0.1132	0.32	0.33	94.8%
$\sigma^2_{\eta_1}$	15.0	15.46	9.3072	2.78	3.02	93%	15.0	15.18	4.2164	1.93	2.05	93.2%
η_2	-0.2	-0.20	0.0021	0.05	0.05	94.2%	-0.2	-0.20	0.0009	0.03	0.03	94.4%
$\sigma^2_{\eta_2}$	0.1	0.10	0.0006	0.02	0.02	93.2%	0.1	0.10	0.0002	0.01	0.01	94.8%
η_3	-3.0	-3.01	0.0320	0.17	0.18	94%	-3.0	-3.01	0.0193	0.13	0.14	95.4%
$\sigma^2_{\eta_3}$	2.0	1.99	0.1181	0.34	0.34	94%	2.0	2.02	0.0711	0.27	0.27	95%
η_4	15.0	15.07	0.3528	0.79	0.59	99.6%	15.0	15.04	0.2782	0.53	0.53	94.8%
$\sigma^2_{\eta_4}$	16.0	16.50	9.9565	3.07	3.12	93%	16.0	16.33	5.4494	2.30	2.31	94%
η_5	28.0	28.01	0.1385	0.37	0.37	95.4%	28.0	28.00	0.0941	0.31	0.31	96.4%
$\sigma^2_{\eta_5}$	16.0	16.27	5.0602	2.29	2.23	94.4%	16.0	16.10	3.2051	1.87	1.79	95.4%
η_6	0.2	0.20	0.0022	0.05	0.05	94.4%	0.2	0.20	0.0013	0.04	0.04	94.6%
$\sigma^2_{\eta_6}$	0.2	0.20	0.0010	0.03	0.03	96%	0.2	0.20	0.0005	0.02	0.02	95.8%
η_7	-0.4	-0.40	0.0015	0.04	0.04	92.6%	-0.4	-0.40	0.0011	0.03	0.03	94%
$\sigma^2_{\eta_7}$	0.2	0.20	0.0008	0.03	0.03	94%	0.2	0.20	0.0004	0.02	0.02	96%
η_8	10.0	10.17	0.3413	0.83	0.56	99.4%	10.0	10.12	0.2775	0.61	0.51	96.2%
$\sigma^2_{\eta_8}$	4.0	4.83	3.2965	1.75	1.62	95.6%	4.0	4.24	0.9806	0.97	0.96	94%
$r_{\eta_4\eta_8}$	0.4	0.34	0.0395	0.20	0.19	95.8%	0.4	0.38	0.0161	0.13	0.13	94%
$\sigma^2_{\epsilon_1}$	20.0	20.06	0.8345	0.89	0.91	94.2%	4.0	4.01	0.0371	0.19	0.19	94.8%
$\sigma^2_{\epsilon_2}$	5.0	5.01	0.0499	0.22	0.22	95.6%	1.0	1.00	0.0022	0.05	0.05	95.4%
ε_1	3.0	2.88	0.3447	1.26	0.58	100%	3.0	2.96	0.6141	0.83	0.78	95%
ε_2	3.0	2.60	0.3967	1.33	0.49	100%	3.0	2.76	0.7212	1.07	0.81	97%
$r_{\eta_2\eta_3}$	0.2	0.23	0.0360	0.18	0.19	92.4%	0.2	0.21	0.0154	0.12	0.12	94.2%
$r_{\eta_6\eta_7}$	-0.5	-0.51	0.0088	0.09	0.09	94.4%	-0.5	-0.50	0.0049	0.07	0.07	93.4%

Table 2.10: Simulation results for comparing three bivariate models under scenarios 1 and 2

	Broken-Stick						Bacan-Watts						Smooth Polynomial					
				Mean	Emp.	95%PI				Mean	Emp.	95%PI				Mean	Emp.	95%PI
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
Scenario 1																		
$\alpha_4(16.5)$	14.29	5.502	0.77	0.78	19%	$\beta_4(16.5)$	14.25	5.666	0.77	0.78	17.8%	$\eta_4(15)$	15.07	1.085	0.99	1.04	93.4%	
$\sigma^2_{\alpha_4}(64)$	48.68	286.562	7.41	7.21	50.4%	$\sigma^2_{\beta_4}(64)$	48.37	294.211	7.31	7.08	48.8%	$\sigma^2_{\eta_4}(64)$	66.11	124.185	11.21	10.95	94.6%	
$\alpha_8(11.5)$	9.16	6.606	0.93	1.07	31.4%	$\beta_8(11.5)$	9.30	5.886	0.91	1.02	33.6%	$\eta_8(10)$	9.87	0.664	0.85	0.81	95.8%	
$\sigma^2_{\alpha_8}(16)$	26.27	148.108	7.43	6.53	67.8%	$\sigma^2_{\beta_8}(16)$	25.59	132.982	7.15	6.41	68.6%	$\sigma^2_{\eta_8}(16)$	20.86	45.994	5.41	4.73	88.6%	
$r_{\alpha_4\alpha_8}(0.2)$	0.12	0.027	0.15	0.14	94.8%	$r_{\beta_4\beta_8}(0.2)$	0.12	0.026	0.15	0.14	94.6%	$r_{\eta_4\eta_8}(0.2)$	0.17	0.022	0.15	0.15	96%	
Scenario 2																		
$\alpha_4(16.5)$	14.18	5.793	0.61	0.63	4.6%	$\beta_4(16.5)$	14.23	5.551	0.61	0.63	5%	$\eta_4(15)$	15.04	0.575	0.73	0.76	95.2%	
$\sigma^2_{\alpha_4}(64)$	49.61	245.138	6.52	6.19	46.2%	$\sigma^2_{\beta_4}(64)$	49.24	256.029	6.48	6.17	42.2%	$\sigma^2_{\eta_4}(64)$	65.53	86.804	9.27	9.20	95.2%	
$\alpha_8(11.5)$	10.72	0.983	0.54	0.61	66.6%	$\beta_8(11.5)$	10.65	1.102	0.55	0.61	65.2%	$\eta_8(10)$	9.95	0.245	0.52	0.49	94.8%	
$\sigma^2_{\alpha_8}(16)$	17.51	20.782	3.62	4.30	89%	$\sigma^2_{\beta_8}(16)$	18.13	23.718	3.82	4.38	88.4%	$\sigma^2_{\eta_8}(16)$	18.23	16.328	3.41	3.38	92%	
$r_{\alpha_4\alpha_8}(0.2)$	0.18	0.016	0.13	0.12	95.8%	$r_{\beta_4\beta_8}(0.2)$	0.18	0.015	0.13	0.12	97%	$r_{\eta_4\eta_8}(0.2)$	0.18	0.014	0.12	0.12	96.2%	

Table 2.11: Simulation results for comparing three bivariate models under scenarios 3 and 4

	Broken-Stick						Bacan-Watts						Smooth Polynomial					
				Mean	Emp.	95%PI				Mean	Emp.	95%PI				Mean	Emp.	95%PI
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
									Scenario 3									
$\alpha_4(16.5)$	15.75	0.796	0.46	0.48	62.2%		$\beta_4(16.5)$	15.76	0.795	0.47	0.49	63.2%	$\eta_4(15)$	15.01	0.276	0.50	0.53	94.4%
$\sigma^2_{\alpha_4}(16)$	16.08	7.266	2.74	2.70	94.2%		$\sigma^2_{\beta_4}(16)$	16.20	7.376	2.74	2.71	94.2%	$\sigma^2_{\eta_4}(16)$	16.33	10.357	3.04	3.20	92.2%
$\alpha_8(11.5)$	11.26	0.435	0.50	0.61	87.4%		$\beta_8(11.5)$	11.24	0.393	0.49	0.57	87.6%	$\eta_8(10)$	10.01	0.229	0.48	0.48	94%
$\sigma^2_{\alpha_8}(4)$	3.28	11.610	1.95	3.33	71.8%		$\sigma^2_{\beta_8}(4)$	2.93	9.159	1.81	2.84	70.6%	$\sigma^2_{\eta_8}(4)$	4.63	3.135	1.73	1.66	96.2%
$r_{\alpha_4\alpha_8}(0.2)$	0.23	0.100	0.30	0.32	91.2%		$r_{\beta_4\beta_8}(0.2)$	0.24	0.098	0.30	0.31	91.4%	$r_{\eta_4\eta_8}(0.2)$	0.18	0.047	0.22	0.22	94.4%
									Scenario 4									
$\alpha_4(16.5)$	15.85	0.548	0.35	0.36	54.4%		$\beta_4(16.5)$	15.87	0.521	0.35	0.36	57.8%	$\eta_4(15)$	15.02	0.133	0.36	0.36	94.2%
$\sigma^2_{\alpha_4}(16)$	15.33	5.039	2.12	2.15	91.6%		$\sigma^2_{\beta_4}(16)$	15.28	5.141	2.11	2.15	92%	$\sigma^2_{\eta_4}(16)$	16.26	5.533	2.30	2.34	93.4%
$\alpha_8(11.5)$	11.45	0.083	0.27	0.28	92.4%		$\beta_8(11.5)$	11.43	0.086	0.27	0.28	92.8%	$\eta_8(10)$	10.02	0.070	0.27	0.26	95.6%
$\sigma^2_{\alpha_8}(4)$	3.79	1.170	0.97	1.06	90.4%		$\sigma^2_{\beta_8}(4)$	3.76	1.300	1.02	1.12	90.8%	$\sigma^2_{\eta_8}(4)$	4.19	0.964	0.98	0.96	95.2%
$r_{\alpha_4\alpha_8}(0.2)$	0.21	0.023	0.15	0.15	94.4%		$r_{\beta_4\beta_8}(0.2)$	0.21	0.024	0.15	0.16	94.4%	$r_{\eta_4\eta_8}(0.2)$	0.19	0.019	0.14	0.14	94%

Table 2.12: Simulation results for comparing three bivariate models under scenarios 5 and 6

	Broken-Stick						Bacan-Watts						Smooth Polynomial					
				Mean	Emp.	95%PI				Mean	Emp.	95%PI				Mean	Emp.	95%PI
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
Scenario 5																		
$\alpha_4(16.5)$	14.33	5.326	0.77	0.78	20.4%	$\beta_4(16.5)$	14.30	5.448	0.77	0.79	20.4%	$\eta_4(15)$	15.08	1.040	0.98	1.02	94.2%	
$\sigma^2_{\alpha_4}(64)$	48.85	282.450	7.43	7.28	52.4%	$\sigma^2_{\beta_4}(64)$	48.61	288.231	7.34	7.19	50.2%	$\sigma^2_{\eta_4}(64)$	66.14	121.533	11.16	10.83	95.4%	
$\alpha_8(11.5)$	9.49	5.144	0.88	1.05	36.4%	$\beta_8(11.5)$	9.63	4.477	0.86	1.00	41.4%	$\eta_8(10)$	9.87	0.648	0.83	0.79	95%	
$\sigma^2_{\alpha_8}(16)$	20.46	73.216	6.75	7.31	88%	$\sigma^2_{\beta_8}(16)$	19.74	65.048	6.50	7.15	89.4%	$\sigma^2_{\eta_8}(16)$	21.03	47.545	5.26	4.72	87.2%	
$r_{\alpha_4\alpha_8}(0.4)$	0.28	0.043	0.17	0.17	89.6%	$r_{\beta_4\beta_8}(0.4)$	0.29	0.041	0.17	0.17	91.8%	$r_{\eta_4\eta_8}(0.4)$	0.33	0.022	0.14	0.13	95%	
Scenario 6																		
$\alpha_4(16.5)$	14.23	5.537	0.61	0.63	4%	$\beta_4(16.5)$	14.28	5.314	0.61	0.63	6%	$\eta_4(15)$	15.06	0.582	0.72	0.76	94.8%	
$\sigma^2_{\alpha_4}(64)$	49.57	246.566	6.49	6.21	45.4%	$\sigma^2_{\beta_4}(64)$	49.27	255.149	6.44	6.18	43%	$\sigma^2_{\eta_4}(64)$	65.57	85.029	9.20	9.10	95.6%	
$\alpha_8(11.5)$	10.78	0.833	0.52	0.56	69.4%	$\beta_8(11.5)$	10.73	0.943	0.54	0.59	66.6%	$\eta_8(10)$	9.95	0.225	0.50	0.47	96.2%	
$\sigma^2_{\alpha_8}(16)$	17.19	18.251	3.53	4.11	91.4%	$\sigma^2_{\beta_8}(16)$	17.64	19.782	3.65	4.14	90.4%	$\sigma^2_{\eta_8}(16)$	18.28	15.958	3.36	3.28	91.2%	
$r_{\alpha_4\alpha_8}(0.4)$	0.33	0.019	0.12	0.12	93.2%	$r_{\beta_4\beta_8}(0.4)$	0.33	0.019	0.12	0.12	91.8%	$r_{\eta_4\eta_8}(0.4)$	0.36	0.012	0.11	0.10	95.6%	

Table 2.13: Simulation results for comparing three bivariate models under scenarios 7 and 8

	Broken-Stick						Bacan-Watts						Smooth Polynomial					
				Mean	Emp.	95%PI				Mean	Emp.	95%PI				Mean	Emp.	95%PI
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
									Scenario 7									
$\alpha_4(16.5)$	15.76	0.785	0.47	0.48	63%	$\beta_4(16.5)$	15.76	0.790	0.47	0.49	63.2%	$\eta_4(15)$	15.01	0.276	0.50	0.53	93.4%	
$\sigma^2_{\alpha_4}(16)$	16.10	7.386	2.74	2.72	94.2%	$\sigma^2_{\beta_4}(16)$	16.21	7.628	2.74	2.76	93.6%	$\sigma^2_{\eta_4}(16)$	16.36	10.377	3.02	3.20	92%	
$\alpha_8(11.5)$	11.27	0.394	0.49	0.59	88%	$\beta_8(11.5)$	11.25	0.383	0.49	0.57	87.2%	$\eta_8(10)$	10.01	0.221	0.47	0.47	95%	
$\sigma^2_{\alpha_8}(4)$	3.31	9.244	1.91	2.96	76.8%	$\sigma^2_{\beta_8}(4)$	3.07	7.688	1.80	2.61	74.6%	$\sigma^2_{\eta_8}(4)$	4.73	3.007	1.70	1.57	96.4%	
$r_{\alpha_4\alpha_8}(0.4)$	0.41	0.072	0.27	0.27	94%	$r_{\beta_4\beta_8}(0.4)$	0.42	0.071	0.26	0.27	94%	$r_{\eta_4\eta_8}(0.4)$	0.35	0.039	0.20	0.19	95.6%	
									Scenario 8									
$\alpha_4(16.5)$	15.87	0.524	0.34	0.36	55.6%	$\beta_4(16.5)$	15.89	0.501	0.34	0.36	58.2%	$\eta_4(15)$	15.02	0.132	0.35	0.36	94.6%	
$\sigma^2_{\alpha_4}(16)$	15.27	5.078	2.10	2.13	91.4%	$\sigma^2_{\beta_4}(16)$	15.22	5.086	2.09	2.12	92.2%	$\sigma^2_{\eta_4}(16)$	16.24	5.473	2.28	2.33	93.8%	
$\alpha_8(11.5)$	11.46	0.079	0.26	0.28	93.6%	$\beta_8(11.5)$	11.44	0.084	0.27	0.28	93%	$\eta_8(10)$	10.02	0.068	0.27	0.26	94.4%	
$\sigma^2_{\alpha_8}(4)$	3.77	1.123	0.94	1.04	89.6%	$\sigma^2_{\beta_8}(4)$	3.70	1.219	0.98	1.06	90.6%	$\sigma^2_{\eta_8}(4)$	4.20	0.932	0.97	0.94	95.6%	
$r_{\alpha_4\alpha_8}(0.4)$	0.40	0.021	0.14	0.14	94.4%	$r_{\beta_4\beta_8}(0.4)$	0.40	0.022	0.14	0.15	93.6%	$r_{\eta_4\eta_8}(0.4)$	0.39	0.016	0.13	0.12	95.8%	

Table 2.14: Simulation results for comparing three bivariate models under scenarios 9 and 10

	Broken-Stick						Bacan-Watts						Smooth Polynomial					
				Mean	Emp. 95%PI					Mean	Emp. 95%PI					Mean	Emp. 95%PI	
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
Scenario 9																		
$\alpha_4(16.5)$	14.35	5.228	0.77	0.78	21.4%	$\beta_4(16.5)$	14.32	5.359	0.77	0.78	19.8%	$\eta_4(15)$	15.12	1.021	0.97	1.00	94%	
$\sigma^2_{\alpha_4}(64)$	48.98	279.853	7.41	7.37	51.8%	$\sigma^2_{\beta_4}(64)$	48.67	286.392	7.30	7.17	49.6%	$\sigma^2_{\eta_4}(64)$	66.51	124.151	11.09	10.87	95%	
$\alpha_8(11.5)$	9.64	4.484	0.85	1.01	39.2%	$\beta_8(11.5)$	9.77	3.891	0.84	0.95	42.6%	$\eta_8(10)$	9.87	0.550	0.79	0.73	95.6%	
$\sigma^2_{\alpha_8}(16)$	19.72	63.617	6.38	7.06	89%	$\sigma^2_{\beta_8}(16)$	19.16	56.178	6.14	6.80	90%	$\sigma^2_{\eta_8}(16)$	21.30	48.324	5.10	4.51	84.6%	
$r_{\alpha_4\alpha_8}(0.6)$	0.42	0.057	0.16	0.16	82.4%	$r_{\beta_4\beta_8}(0.6)$	0.43	0.052	0.16	0.16	83%	$r_{\eta_4\eta_8}(0.6)$	0.48	0.026	0.12	0.11	88.8%	
Scenario 10																		
$\alpha_4(16.5)$	14.32	5.132	0.60	0.63	5.6%	$\beta_4(16.5)$	14.38	4.909	0.60	0.64	7%	$\eta_4(15)$	15.09	0.557	0.71	0.74	94.8%	
$\sigma^2_{\alpha_4}(64)$	49.67	244.287	6.45	6.24	45%	$\sigma^2_{\beta_4}(64)$	49.38	251.361	6.41	6.15	42%	$\sigma^2_{\eta_4}(64)$	65.78	84.144	9.11	9.01	95.6%	
$\alpha_8(11.5)$	10.84	0.730	0.50	0.54	70.4%	$\beta_8(11.5)$	10.81	0.772	0.51	0.55	70.6%	$\eta_8(10)$	9.95	0.201	0.47	0.45	96.6%	
$\sigma^2_{\alpha_8}(16)$	16.75	15.109	3.35	3.82	93.8%	$\sigma^2_{\beta_8}(16)$	17.07	16.824	3.45	3.96	91.2%	$\sigma^2_{\eta_8}(16)$	18.38	15.352	3.26	3.11	91.2%	
$r_{\alpha_4\alpha_8}(0.6)$	0.51	0.020	0.10	0.11	86.8%	$r_{\beta_4\beta_8}(0.6)$	0.51	0.020	0.10	0.11	86.8%	$r_{\eta_4\eta_8}(0.6)$	0.54	0.011	0.09	0.08	91.4%	

Table 2.15: Simulation results for comparing three bivariate models under scenarios 11 and 12

	Broken-Stick						Bacan-Watts						Smooth Polynomial					
				Mean	Emp.	95%PI				Mean	Emp.	95%PI				Mean	Emp.	95%PI
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
Scenario 11																		
$\alpha_4(16.5)$	15.77	0.766		0.46	0.48	64.4%	$\beta_4(16.5)$	15.77	0.764	0.47	0.48	66%	$\eta_4(15)$	15.02	0.264	0.49	0.51	94.2%
$\sigma^2_{\alpha_4}(16)$	16.09	7.309		2.72	2.70	94.2%	$\sigma^2_{\beta_4}(16)$	16.22	7.475	2.73	2.73	94.2%	$\sigma^2_{\eta_4}(16)$	16.40	10.128	3.02	3.16	92.4%
$\alpha_8(11.5)$	11.30	0.361		0.48	0.57	88.8%	$\beta_8(11.5)$	11.27	0.336	0.48	0.53	89.8%	$\eta_8(10)$	10.01	0.215	0.47	0.46	96%
$\sigma^2_{\alpha_8}(4)$	3.44	7.772		1.83	2.73	82.6%	$\sigma^2_{\beta_8}(4)$	3.20	6.053	1.77	2.33	83%	$\sigma^2_{\eta_8}(4)$	4.93	3.169	1.70	1.52	95.6%
$r_{\alpha_4\alpha_8}(0.6)$	0.57	0.045		0.22	0.21	96%	$r_{\beta_4\beta_8}(0.6)$	0.58	0.043	0.22	0.21	97.8%	$r_{\eta_4\eta_8}(0.6)$	0.50	0.034	0.17	0.16	95.8%
Scenario 12																		
$\alpha_4(16.5)$	15.90	0.481		0.34	0.35	59.2%	$\beta_4(16.5)$	15.92	0.463	0.34	0.35	59.6%	$\eta_4(15)$	15.02	0.129	0.35	0.36	94%
$\sigma^2_{\alpha_4}(16)$	15.20	5.042		2.07	2.10	91.2%	$\sigma^2_{\beta_4}(16)$	15.16	5.166	2.07	2.11	91.6%	$\sigma^2_{\eta_4}(16)$	16.25	5.317	2.26	2.29	93.6%
$\alpha_8(11.5)$	11.46	0.075		0.25	0.27	94.2%	$\beta_8(11.5)$	11.45	0.079	0.26	0.28	93.4%	$\eta_8(10)$	10.02	0.064	0.26	0.25	95.8%
$\sigma^2_{\alpha_8}(4)$	3.76	1.007		0.91	0.98	91.4%	$\sigma^2_{\beta_8}(4)$	3.70	1.133	0.94	1.02	90.4%	$\sigma^2_{\eta_8}(4)$	4.28	0.908	0.94	0.91	94.8%
$r_{\alpha_4\alpha_8}(0.6)$	0.59	0.016		0.12	0.13	93.4%	$r_{\beta_4\beta_8}(0.6)$	0.60	0.016	0.12	0.13	92.8%	$r_{\eta_4\eta_8}(0.6)$	0.57	0.011	0.11	0.10	95.6%

Table 2.16: Simulation results for comparing three bivariate models with data generated from a bivariate random smooth polynomial model using lognormal distribution for all random effects and errors.

	broken-stick						Bacon-Watts						Smooth Polynomial					
				Mean	Emp.	95%PI				Mean	Emp.	95%PI				Mean	Emp.	95%PI
	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.	True	Mean	MSE	SD	SD	Cov.
Scenario 1 (large Variances)																		
	$\alpha_4(16.5)$	13.0	13.12	0.51	0.82	12%	$\beta_4(16.5)$	13.0	13.23	0.51	0.82	11%	$\eta_4(15)$	12.9	6.23	0.66	1.30	94%
	$\sigma^2_{\alpha_4}(64)$	20.7	1890	3.37	4.27	35%	$\sigma^2_{\beta_4}(64)$	20.6	1899	3.33	4.19	35%	$\sigma^2_{\eta_4}(64)$	27.0	1443	4.96	8.53	95%
	$\alpha_8(11.5)$	9.9	6.96	0.75	2.10	0.2%	$\beta_8(11.5)$	9.9	6.78	0.75	2.08	0.4%	$\eta_8(10)$	7.6	14.07	0.84	2.91	95%
	$\sigma^2_{\alpha_8}(16)$	19.2	524	5.56	22.70	67%	$\sigma^2_{\beta_8}(16)$	18.3	429	5.36	20.62	67%	$\sigma^2_{\eta_8}(16)$	16.8	102	4.00	10.11	84%
	$r_{\alpha_4\alpha_8}(0.4)$	0.3	0.05	0.17	0.16	75%	$r_{\beta_4\beta_8}(0.4)$	0.3	0.05	0.17	0.16	75%	$r_{\eta_4\eta_8}(0.4)$	0.3	0.04	0.15	0.15	93%
Scenario 2 (small Variances)																		
	$\alpha_4(16.5)$	15.5	1.08	0.29	0.35	2%	$\beta_4(16.5)$	15.5	1.08	0.29	0.35	3%	$\eta_4(15)$	14.5	0.42	0.30	0.39	94%
	$\sigma^2_{\alpha_4}(16)$	11.4	24.36	1.56	1.79	41%	$\sigma^2_{\beta_4}(16)$	11.4	24.65	1.56	1.79	42%	$\sigma^2_{\eta_4}(16)$	12.3	18.16	1.75	2.15	93%
	$\alpha_8(11.5)$	11.1	2.12	0.32	1.40	7%	$\beta_8(11.5)$	11.0	2.20	0.33	1.40	9%	$\eta_8(10)$	9.4	0.60	0.31	0.44	95%
	$\sigma^2_{\alpha_8}(4)$	5.2	80.51	1.29	8.90	47%	$\sigma^2_{\beta_8}(4)$	4.8	40.72	1.20	6.34	52%	$\sigma^2_{\eta_8}(4)$	4.6	2.00	1.04	1.30	90%
	$r_{\alpha_4\alpha_8}(0.4)$	0.4	0.03	0.15	0.16	82%	$r_{\beta_4\beta_8}(0.4)$	0.4	0.03	0.15	0.17	81%	$r_{\eta_4\eta_8}(0.4)$	0.4	0.02	0.14	0.14	96%

2.6 Application to the IIDS Data

In this section, the three proposed models are fitted to the IIDS data using the previously described Bayesian method. Specifically, age was centered at 65 years. For each bivariate model, we consider two different models that differ in the variance-covariance structure. Of particular interest is the relationship between the change points of cognitive and BMI measurements, and we assume that cognitive function and BMI are correlated only through their change points.

For bivariate random broken-stick models, we denote by BS_1 the model with the following specified variance-covariance structure:

$$\mathbf{\Sigma}_\alpha = \begin{pmatrix} \sigma^2_{\alpha_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\alpha_2} & \sigma_{\alpha_2\alpha_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{\alpha_2\alpha_3} & \sigma^2_{\alpha_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\alpha_4} & 0 & 0 & 0 & \sigma_{\alpha_4\alpha_8} \\ 0 & 0 & 0 & 0 & \sigma^2_{\alpha_5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\alpha_6} & \sigma_{\alpha_6\alpha_7} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\alpha_6\alpha_7} & \sigma^2_{\alpha_7} & 0 \\ 0 & 0 & 0 & \sigma_{\alpha_4\alpha_8} & 0 & 0 & 0 & \sigma^2_{\alpha_8} \end{pmatrix}.$$

This model allows correlations between the slopes before and after the change points. Model BS_2 is further defined by setting $\sigma_{\alpha_2\alpha_3} = 0$ and $\sigma_{\alpha_6\alpha_7} = 0$, without correlations between the two slopes. Bayesian estimation for the above two models were obtained by imposing prior distributions similar to those in the simulation study. We found that the choice of non-informative prior distributions had little influence on the marginal posterior distributions.

Model BS_1 (DIC $= 15,810$, LPML $= -6,752$) is superior to BS_2 (DIC $= 15,820$, LPML $= -6,766$) in terms of smaller DIC, larger LPML, and better convergence based on the history trace plots of model parameters. Table 4 shows the summary statistics of the posterior distributions of model parameters from model BS_1. The mean (95% posterior interval) age of cognitive function change point is 22.7 $(19.5, 25.7)$ years; the mean slope before change point is -0.1 $(-0.2, -0.1)$ points/year, and the mean slope after change point is -1.7 $(-2.6, -1.0)$ points/year. Cognitive function decreases steadily before the change point and plummets after the change point. On the other hand, for BMI the mean age of change point is 12.5 $(9.0, 16.8)$ years; the mean slope before the change point is 0.2 $(0.03, 0.4)$ points/year, and the mean slope after the change point is -0.4 $(-0.6, -0.3)$ points/year. BMI steadily increases before its change point and decreases after. The posterior means of correlation between the 2 slopes for cognitive function and BMI are both positive but with wide 95% posterior intervals. The posterior mean of the correlation between the 2 change points is 0.5 $(0.1, 0.8)$, suggesting that the change in cognitive function is positively correlated with the change in BMI. Furthermore, the estimated change point of BMI is found to be on average 10 years ahead of the estimated change point of cognitive function. The posterior means of the variances of the change points of cognitive function and BMI are 52.3 $(30.5, 82.8)$ and 22.5 $(9.0, 43.6)$, respectively.

For the bivariate random Bacon-Watts model, we denote by BW_1 the model with the same variance-covariance structure as for BS_1, and by BW_2 the model with the same variance-covariance structure as for BS_2 above. Prior distributions are chosen to be similar to the settings for the bivariate random Bacon-Watts model in the simulation study. Different prior distributions led to similar posterior distributions of model parameters, DIC and LPML. The two slopes in each model are correlated according to the model parameterization of the random Bacon-Watts model. Therefore BW_1 as a more faithful model should

be a better model than BW_2. This is evident from the history trace plots of model parameters, the DIC, and LPML. Specifically, the BW_1 had a smaller DIC $15,790$ vs. $15,870$ from model BW_2, and a greater LPML $-6,738$ vs. $-6,806$ from model BW_2. The posterior distributions of model parameters for model BW_1 is summarized in Table 3. The mean change points for cognitive function and the BMI are 22.1 $(18.8, 25.5)$ and 11.3 $(7.9, 15.3)$ years, respectively. The posterior mean of transition parameters ϕ_1 and ϕ_2 are 1.7 $(0.2, 4.4)$ and 3.3 $(0.4, 4.9)$, respectively. The posterior mean of change point correlation is 0.5 $(0.1, 0.8)$, confirming that change in cognitive function is positively correlated with change in BMI. Similar to the previous model BS_1, the estimated change point of BMI is around 11 years ahead of the estimated change point of cognitive function.

Again, for the bivariate random smooth polynomial model, model SP_1 enjoys the same variance-covariance structure as for BS_1, and model SP_2 has the same variance-covariance structure as for BS_2. The prior distributions for parameters in the above two models are specified similarly as in the simulation study. Different from the bivariate random Bacon-Watts model, the transition parameters in the smooth polynomial model are held constant at 3 years, according to the roughly 3 year intervals between two visits. We assume that the change of cognitive function and BMI occurred within 3 years (we also implemented model fitting with the two transition parameters equal to 1 or 6 years, but parameter estimates are quite similar). Due to the limited number of repeated measurements per subject, we choose not to include those parameters in the model. This also resulted in a convergence problem in the Bayesian computation. Model SP_1 is superior with a smaller DIC of $15,540$ (compared to Model $SP_2's$ $15,720$), and a larger LPML of $-6,659$ (compared to Model $SP_2's$ $-6,742$), as well as better convergence profiles, as is evident from the history trace plot. Table 5 shows a summary of the posterior distributions of model parameters for model SP_1. The mean age of cognitive function change point is 27.5 $(= 26.0 + 1.5)$ $(22.3, 30.8)$ years

with variance 60.8 $(33.5, 105.6)$; the mean slope before smooth interval is -0.2 $(-0.2, -0.1)$ points/year, and the mean slope after smooth interval is -2.99 $(-4.8, -1.67)$ points/year. Cognitive function steadily decreases before the smooth interval and sharply declines after. For BMI measurement, the mean age of BMI change point is 11.0 $(= 9.5 + 1.5)$ $(7.3, 12.4)$ years with variance 14.5 $(8.0, 23.9)$; the mean slope before the smooth interval is 0.2 $(0.1, 0.4)$ points/year, and the mean slope after the smooth interval is -0.4 $(-0.5, -0.3)$ points/year. The BMI has a similar trend - gradually increasing before smooth interval and decreasing after. The posterior mean of change point correlation is 0.6 $(0.2, 0.8)$, which again implies that change in cognitive function is positively correlated with the change in BMI. Compared with the change point of cognitive function, the change point of BMI appears to be 16 years ahead on average.

The model fitting of the three bivariate models uses the same 10,000 burn-in and 40,000 additional iterations. It takes about 20 minutes, 27 minutes, and 55 minutes for the bivariate random broken-stick, the bivariate random Bacon-Watts and the bivariate random smooth polynomial model, respectively, in a PC with Pentium(R) 4 CPU 2.00 GB of RAM.

In the previous paragraphs, we have presented model-fitting results of the 3 different bivariate random change point models for cognitive function and BMI. The fitted trajectories of nine random selected individuals from IIDS for models BS1, BW1, and SP1 are shown in Figure 2.3. The three models were compared based on both DIC and LPML; model SP_1 appears to be the best model under consideration of the smallest DIC $(15,540)$ and the largest LPML $(-6,659)$. Differences in parameter estimation among the three models were observed (Table 2.17). The estimated change points of cognitive function measurements are $22.7, 22.1$, and 27.5 years for model BS_1, BW_1, and SP_1, respectively. The estimated change points of models BS_1 and BW_1 are very close, while the estimated change point of cognitive function of SP_1 is around 5 years later than those from the other two models.

Various reasons may cause such a big change point estimate for cognitive function in the random smooth polynomial model. First, as we have observed in scenario 1 of the simulation study, if the true model is the bivariate random smooth polynomial model with larger variances, the estimated change points tends to be years later than the other two models. Secondly, the special model structure of the smooth polynomial model (with an additional smooth interval between two linear trends) allows the seeking of change points in a later time window. This is observed in the individual trajectory plot as well. On the other hand, estimated change points for BMI in three models are comparable, which are all around 12 years ($12.5, 11.3$, and 11.0 years for model BS_1, BW_1, and SP_1, respectively). This may be because the BMI increases first and decreases later making it easy to detect a change point for change point models. Another possibility is that the variance of measurement error and variance of change point are both much smaller compared to the cognitive function, so the estimation of change point of BMI becomes more stable for different models. The estimated correlation between the change points of cognitive function and BMI in the 3 joint models are similar($r_{\alpha_4\alpha_8} = 0.5$, $r_{\beta_4\beta_8} = 0.5$ and $r_{\eta_4\eta_8} = 0.6$), and all of them have good 95% posterior interval coverage.

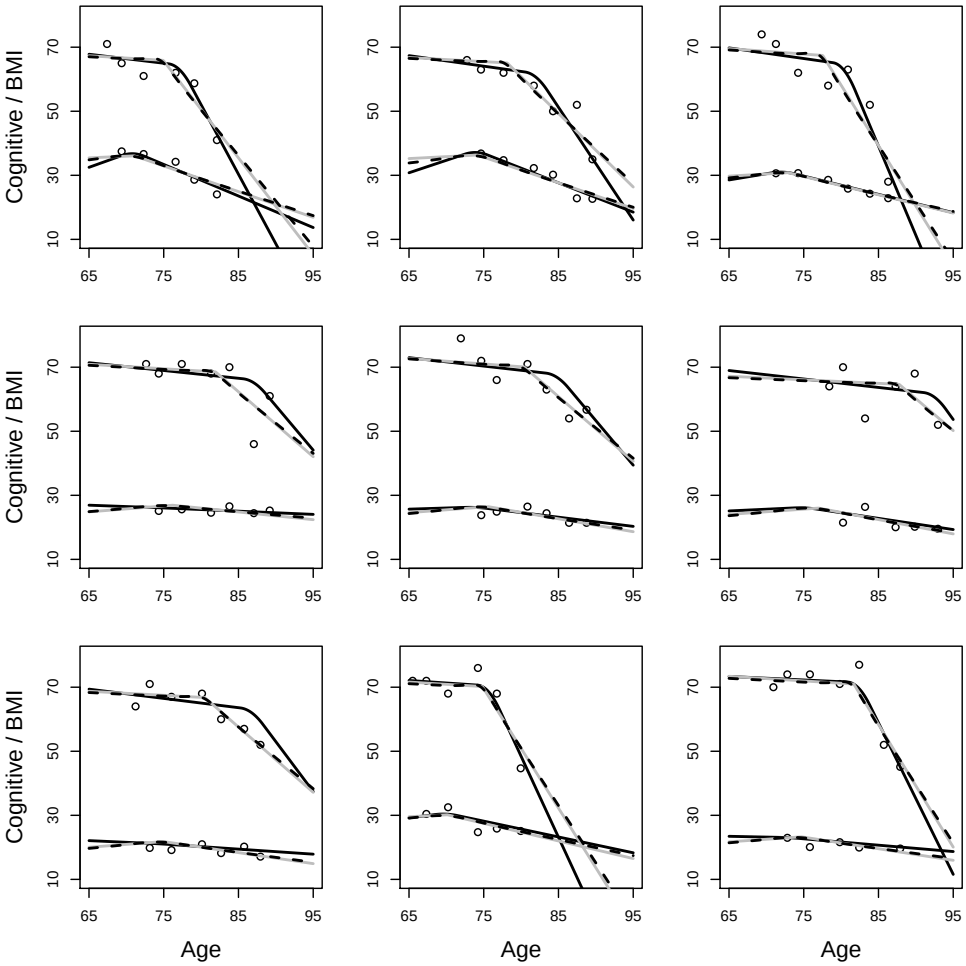

Figure 2.3: Plots of nine random selected participants from IIDS (black circle), fit for bivariate random broken-stick model BS_1 (solid gray line), bivariate random Bacon-Watts model BW_1 (dashed black line) and bivariate random smooth polynomial model SP_1 (solid black line). The three fitted curves on the top are for cognitive scores, and the three fitted curves on the bottom are for BMI measures.

Table 2.17: Bayesian estimates of population parameters and 95% Posterior Interval (95% PI) for bivariate random broken-stick model (BS_1), bivariate random Bacon-Watts model (BW_1) and bivariate random smooth polynomial model (SP_1) from IIDS data.

broken-stick model BS_1			Bacon-Watts model BW_1			smooth polynomial model SP_1		
Para.	Est.	95% PI	Para.	Est.	95% PI	Para.	Est.	95% PI
α_1	68.6	(67.4, 69.7)	β_1	68.5	(67.4, 69.7)	η_1	71.4	(70.6, 72.1)
$\sigma^2_{\alpha_1}$	14.8	(10.8, 19.4)	$\sigma^2_{\beta_1}$	14.7	(10.7, 19.3)	$\sigma^2_{\eta_1}$	15.1	(11.3, 19.5)
α_2	-0.1	(-0.2, -0.1)	β_2	-0.9	(-1.3, -0.5)	η_2	-0.2	(-0.2, -0.1)
$\sigma^2_{\alpha_2}$	0.02	(0.01, 0.03)	$\sigma^2_{\beta_2}$	0.4	(0.1, 1.0)	$\sigma^2_{\eta_2}$	0.04	(0.02, 0.1)
α_3	-1.7	(-2.6, -1.0)	β_3	-0.7	(-1.2, -0.4)	η_3	-3.0	(-4.8, -1.7)
$\sigma^2_{\alpha_3}$	1.8	(0.6, 3.9)	$\sigma^2_{\beta_3}$	0.4	(0.12, 1.0)	$\sigma^2_{\eta_3}$	3.5	(0.9, 8.1)
α_4	22.7	(19.5, 25.7)	β_4	22.1	(18.8, 25.5)	η_4	26.0	(22.3, 30.8)
						$\eta_4 + 1/2\varepsilon_1$	27.5	
$\sigma^2_{\alpha_4}$	52.3	(30.5, 82.8)	$\sigma^2_{\beta_4}$	51.4	(27.8, 83.0)	$\sigma^2_{\eta_4}$	60.8	(33.5, 105.6)
α_5	30.8	(29.9, 31.7)	β_5	30.7	(29.8, 31.6)	η_5	28.6	(27.7, 29.5)
$\sigma^2_{\alpha_5}$	25.9	(21.1, 31.7)	$\sigma^2_{\beta_5}$	26.1	(21.2, 32.0)	$\sigma^2_{\eta_5}$	15.9	(11.9, 20.6)
α_6	0.2	(0.03, 0.4)	β_6	-0.9	(-0.2, 0.1)	η_6	0.2	(0.1, 0.4)
$\sigma^2_{\alpha_6}$	0.03	(0.01, 0.1)	$\sigma^2_{\beta_6}$	0.03	(0.01, 0.05)	$\sigma^2_{\eta_6}$	0.2	(0.1, 0.3)
α_7	-0.4	(-0.6, -0.3)	β_7	-0.3	(-0.4, -0.2)	η_7	-0.4	(-0.5, -0.3)
$\sigma^2_{\alpha_7}$	0.1	(0.04, 0.1)	$\sigma^2_{\beta_7}$	0.04	(0.01, 0.1)	$\sigma^2_{\eta_7}$	0.1	(0.1, 0.2)
α_8	12.5	(9.0, 16.8)	β_8	11.3	(7.9, 15.3)	η_8	9.5	(7.3, 12.4)
						$\eta_8 + 1/2\varepsilon_2$	11.0	
$\sigma^2_{\alpha_8}$	22.5	(9.0, 43.6)	$\sigma^2_{\beta_8}$	18.4	(8.5, 36.1)	$\sigma^2_{\eta_8}$	14.5	(8.0, 23.9)
$r_{\alpha_4\alpha_8}$	0.5	(0.1, 0.8)	$r_{\beta_4\beta_8}$	0.5	(0.1, 0.8)	$r_{\eta_4\eta_8}$	0.6	(0.2, 0.8)
$\sigma^2_{\epsilon_1}$	20.1	(18.4, 21.9)	$\sigma^2_{\epsilon_1}$	20.0	(18.3, 21.9)	$\sigma^2_{\epsilon_1}$	20.0	(18.3, 21.7)
$\sigma^2_{\epsilon_2}$	5.8	(5.2, 6.4)	$\sigma^2_{\epsilon_2}$	5.7	(5.0, 6.4)	$\sigma^2_{\epsilon_2}$	5.0	(4.5, 5.6)
$r_{\alpha_2\alpha_3}$	0.04	(-0.6, 0.6)	$r_{\beta_2\beta_3}$	1.0	(0.9, 1.0)	$r_{\eta_2\eta_3}$	0.1	(-0.7, 0.8)
$r_{\alpha_6\alpha_7}$	0.1	(-0.65, 0.7)	$r_{\beta_6\beta_7}$	0.2	(-0.5, 0.7)	$r_{\eta_6\eta_7}$	-0.9	(-1.0, -0.8)
			Φ_1	1.7	(0.2, 4.4)			
			Φ_2	3.3	(0.4, 4.9)			

2.7 Conclusion

We have developed joint modeling frameworks of bivariate longitudinal outcomes under three different bivariate random change point models: the bivariate random broken-stick model, the bivariate random Bacon-Watts model, and the bivariate random smooth polynomial model. The proposed methodology was applied to the IIDS data. The Bayesian method was used for model fitting using the *BRugs* package in **R**. The goodness of model fitting was assessed using DIC and LPML.

The Bayesian method has been a useful tool for parameter estimation of mixed-effects models with several advantages compared to traditional frequentist methods. First, the highly complex model structure can be still easily handled in WinBUGS and BRugs. Second, the Bayesian method can deal with the mixed-effect model with multiple random effects. The maximum likelihood method using Gaussian quadrature is commonly used for parameter estimation in non-linear mixed-effect models, but computation of multi-fold integrations can become intractable with large number of random effects. Third, the Bayesian method is also advantageous in its interpretability and ability to deal with missing data (van den Hout et al., 2010). Finally, from a practical implementation perspective, the Bayesian method using an MCMC sampling method is conveniently available in various popular statistical software, such as WinBUGS and *BRugs* in **R**. The often cited disadvantage of heavy computation overhead of Bayesian MCMC has become less an issue with the rapid advances of modern computing technology. Bayesian methods require one to specify the prior distributions, which sometimes may be challenging. In many cases, knowledge of priors of parameters is either unknown, or even non-existent, which makes it very difficult to specify a unique prior distribution. Careful sensitivity analysis is needed to assess the influence of different priors on the posterior estimates. In our analysis, it appears that the model-fitting results are not sensitive to the choices of priors. On the other hand, the

Bayesian method is a useful technique that can incorporate available prior knowledge of the model parameters into the prior distributions.

The restriction of subjects with at least 5 measurements to be included in our analysis makes the models conditional on the subjects having to survive to a relatively long period of time during follow-up and inevitably limits the modeling to a subset of healthier individuals than the rest of the cohort. It is known that missing data, especially under the non-ignorable missing data mechanism, could significantly impact model results. In the IIDS data and in most longitudinal studies involving elderly subjects, subject dropouts due to death account for the majority of missing data. Our current proposed method is limited in its capability to deal with non-ignorable missing data and also in the ability to detect potential change point, followed by rapid death. Ghosh et al. recently studied the effect of informative dropouts in longitudinal outcomes with multiple change points (Ghosh et al., 2010). It will be an important future research topic to study bivariate change point models that incorporate informative dropouts so that inference on the entire longitudinal cohort can be made. Another interesting and important future research is to take censored change points into account in the bivariate change point model.

The proposed bivariate random change point models not only estimate the change points of bivariate longitudinal outcomes, but also investigate the correlation between the change points. The bivariate random broken-stick model has the advantages of easy implementation and interpretable parameter estimation but the non-continuity at change points may result in numeric problems. The bivariate random Bacon-Watts model solves the problem of non-continuity at change points but the parameter estimation loses meaningful interpretations. The bivariate random smooth polynomial model ensures the continuity at change points and meaningful parameter interpretations, but at a cost of more complex model structures. These methodologies are useful for disease prognosis using biomarkers in medi-

cal science, and it also possesses flexibility in model fitting. Although in this paper we have focused on investigating the correlation between the change points, one can readily extend to more complex models by specifying and estimating other correlation parameters such as correlation between the two slopes before the change point as well as the two slopes after the change point in the bivariate model. The extension to multivariate change point models for multiple longitudinal outcomes is also applicable.

2.8 Acknowledgement

The research is supported by National Institutes of Health Grants R01 AG019181, R01 AG09956, and P30 AG10133.

Chapter 3

Joint Models for Multiple Longitudinal Processes and Time-to-event Outcome

3.1 Abstract

Joint models are useful tools to study the association between time-to-event and longitudinal outcomes. Common estimation methods for joint models include two-stage, Bayesian and maximum-likelihood methods. In this work, we extend existing methods and develop a maximum-likelihood estimation method using the expectation-maximization (EM) algorithm for joint models of a time-to-event outcome and multiple longitudinal processes. We assess the performance of the proposed method via simulations and apply the methodology to a data set to assess the association between longitudinal systolic and diastolic blood pressure (BP) measures and time to coronary artery disease (CAD).

3.2 Introduction

Prospective cohort studies or clinical trials with time-to-event as the primary outcome usually collect many longitudinal variables. Longitudinal studies of Alzheimer's disease, for example also collect repeated measures of height, weight, BP measures and many other variables in order to determine disease etiology (Yang and Gao, 2012). Clinical trials on cardiovascular diseases routinely monitor BP measures at regular intervals to ensure patient safety (Rothwell et al., 2010). In addition, the increasing use of electronic medical records (EMR) in many health care systems makes the collection of many longitudinal laboratory measures as well as time to medical events automatic and straightforward. Separate modeling of the longitudinal processes and the survival outcome may not fully discover potential

disease mechanisms. Appropriate statistical methods are needed to utilize the richness of these data in order to identify potential relationships between the longitudinal measures and disease risk.

Before the introduction of joint models, routine statistical practices in epidemiologic research mostly adopted the Cox model (Cox, 1972) using baseline exposure measures. Such an approach implicitly assumes that the exposure variables stay constant over the length of the study, which is unlikely to be true in studies over an extensive period of time. Cox model with observed longitudinal measures as time-dependent covariates (Andersen et al., 1993; Andersen and Gill, 1982; Fleming and Harrington, 1991) incorporates changes in exposure levels over the follow-up period. However, this model assume that the longitudinal outcomes are continuously measured without errors. This assumption may not be realistic because the longitudinal measures are usually intermittently collected. Furthermore, measurement errors in the longitudinal measurements were not considered in this modeling framework. Lastly, the time-dependent Cox model lacks the flexibility to use various functional forms of the underlying longitudinal processes.

To overcome these difficulties, joint models of longitudinal and survival outcomes were proposed by Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997). They used a linear growth curve model for the longitudinal process and a Cox model with the current value of the longitudinal process as time-dependent covariate. Many extensions to these earlier joint models have been proposed. Henderson et al. (2000) modeled the hazard as a function of the history and rate of change of a biomarker. Brown et al. (2005) extended the linear growth curve model to flexible non-parametric subject-specific random-effects models. Yu et al. (2004) considered a survival-cure model for the time-to-event outcome. Huang et al. (2011) and Elashoff et al. (2008) extended the Cox model to competing risks models. Elashoff et al. (2006) extended the joint models from a single survival outcome

to multiple survival outcomes. Njeru Njagi et al. (2013) considered combining conjugate and normal random effects of longitudinal and time-to-event outcomes in joint models to improve model fit. Qiu et al. (2013) considered a generalized linear mixed model for the longitudinal outcome and a discrete survival model with frailty to predict event probabilities. Comprehensive reviews of joint models have been published (Proust-Lima et al., 2012; Sousa, 2011; Tsiatis and Davidian, 2004; Yu et al., 2004).

When multiple longitudinal measures are available, extension to the joint model framework needs to appropriately account for potential correlations among the longitudinal measures. Simultaneous modeling of multiple longitudinal outcomes in joint models offers a number of advantages over separate modeling of each longitudinal outcome (Brown et al., 2005; Elashoff et al., 2006; Rizopoulos and Ghosh, 2011; Song et al., 2002; Xu and Zeger, 2001). First, for correlated longitudinal outcomes it is more relevant to estimate the adjusted association of each longitudinal outcome with the event risk (Rizopoulos and Ghosh, 2011). Second, Fieuws et al. showed that accounting for the correlation between longitudinal measures may substantially enhance the predictive ability of joint models (Fieuws et al., 2008). In addition, two studies found that joint models of multiple longitudinal outcomes are more efficient compared with separate modeling of each outcome in some settings (Gueorguiva and Sanacora, 2006; McCulloch, 2008).

There are three general types of estimation methods in joint models of longitudinal and survival outcomes: the two-stage approach, Bayesian Markov Chain Monte Carlo (MCMC) method, and maximum-likelihood approach. In the two-stage approach, parameter estimation is conducted separately for the longitudinal model and the survival model. Specifically, at the first stage, parameter estimates and predictions are obtained from the longitudinal models without consideration of the survival outcomes. At the second stage, the predicted longitudinal values are used as true exposure levels in a time-dependent Cox model.

Although the two-stage approach is computationally simple, it can incur bias and loss of efficiency by ignoring the time-to-event information when modeling the longitudinal process Albert and Shih (2010); Faucett and Thomas (1996); Sweeting and Thompson (2011), as the survival process in this setting essentially produces non-ignorable missing data for the longitudinal outcomes. The two-stage approach has been discussed by many authors (Albert and Shih, 2010; Dafni and Tsiatis, 1998; Self and Pawitan, 1992; Sweeting and Thompson, 2011; Tsiatis et al., 1995; Ye et al., 2006). Alternatively, both the Bayesian MCMC approach and the maximum-likelihood approach incorporate both types of outcomes into a joint likelihood function and simultaneously estimate model parameters. The Bayesian MCMC approach has been used for joint models of multiple longitudinal and time-to-event outcomes (Brown and Ibrahim, 2003; Brown et al., 2005; Elashoff et al., 2006; He and Luo, 2013; Rizopoulos and Ghosh, 2011). To the best of our knowledge, the maximum-likelihood method has only been applied to the joint models with a single longitudinal outcome (Huang et al., 2011; Rizopoulos, 2012a; Tseng et al., 2005; Tsiatis and Davidian, 2004; Wulfsohn and Tsiatis, 1997). In particular, Rizopoulos developed an R package (JM) using the EM algorithm (Dempster et al., 1977) for joint models of a time-to-event outcome and a single longitudinal outcome (Rizopoulos, 2010, 2012b).

In this work, we develop a maximum-likelihood approach using the EM algorithm for parameter estimation in joint models of multiple longitudinal processes and a time-to-event outcome. Commonly used for maximum-likelihood estimation, the EM algorithm offers computational advantages over direct likelihood maximization (Couvreur, Couvreur) especially in complex likelihood functions involving random effects. The algorithm increases the likelihood function as iteration continues, ensuring numerical stability. Additional efficiency can be gained when some parameters have closed-form solutions in the M-step.

Finally, predicted values are calculated as part of the E-step reducing the need for further computation.

The remainder of this chapter is organized as follows: Section 3.3 describes a primary care patient cohort, a motivating example. Section 3.4 covers the joint models as well as the joint likelihood function. The EM algorithm estimation method and asymptotic inferences of maximum-likelihood parameter estimates are described in Section 3.5. Section 3.6 reports the simulation studies. In Section 3.7 the proposed method is illustrated using a primary care patient cohort data. Finally the chapter is concluded with a discussion in Section 3.8.

3.3 A Primary Care Patient Cohort

A primary care patient cohort was assembled in 1991 as part of depression screening in primary care clinics in Wishard Health Service. From 1991 to 1993, patients age 60 years or older in the Wishard Health Service were consented for depression screening during their regular clinical visits to their primary care physicians. A total of 4,413 primary care patients were initially contacted, of whom 115 refused; 57 were not eligible due to severe cognitive impairment; 284 were not eligible because they were non-English speaking, in prison, in a nursing home, or had a hearing impairment; 3,957 patients were enrolled in the study. Details about the study have been published in Callahan et al. (1994) and Callahan et al. (1994).

Complete EMR data are available for all enrolled patients and the information includes diagnosis of medical conditions, BP measures, laboratory test measures and medications order and dispensing. One of the research interests using data from this cohort is to examine new risk factors for coronary artery disease (CAD) in elderly population. It is well known from the results of prospective cohort studies that high baseline BP is a risk factor for CAD in middle-aged populations (Anderson et al., 1991; Stamler et al., 1993; Wilson et al., 1998),

but few studies have determined the relationship between longitudinal BP measures and the risk of CAD. Even fewer focused on the elderly population who have declining BP with increasing age. Therefore, it is necessary to apply joint models to determine the association between the longitudinal BP measures and risk of CAD in this elderly cohort.

Among the 3,957 patients enrolled, 2,654 (797 males and 1857 females) were free of CAD at enrollment. For patients with incident CAD events, the date of diagnosis was used as the event time; for patients without CAD, the last outpatient clinic visit before December 31, 2010 was used as the censoring time. Systolic and diastolic BP measured in sitting position from outpatient clinic visits were also collected during follow-up for up to 20 years. Since it has been shown that males have significantly increased CAD risk than females (Hochman et al., 1999; Vaccarino et al., 1999), we focus our analysis on the 797 male patients in the cohort where 28% had incident CAD during the follow-up period from enrollment to December 31, 2010. Mean age of patients included in the analysis sample at baseline was 68 (SD=7.4) years, 519 (65.1%) were black, 254 (31.9%) were smokers, and 268 (33.6%) had history of diabetes at baseline. The frequency of BP measurements varied from patient to patient with a mean frequency of 20.5 (SD=20). For computational convenience annualized systolic and diastolic BP measures during the study period were derived for each participant. On average, there were about 5.3 (SD=4.4) BP measures per subject. Figure 3.1 plots the annualized longitudinal systolic and diastolic BP measures over time by CAD status. The blue and green curves represent fitted population mean BP profiles for CAD and non-CAD groups respectively, using linear mixed-effects models with fixed quadratic time effect. It can be seen that the population mean systolic and diastolic BP measures were higher over time for the CAD group than that for the non-CAD group, indicating a potential association between the risk of CAD and longitudinal systolic and diastolic BP measures. The figure also shows that the differences in BP measures at baseline between the

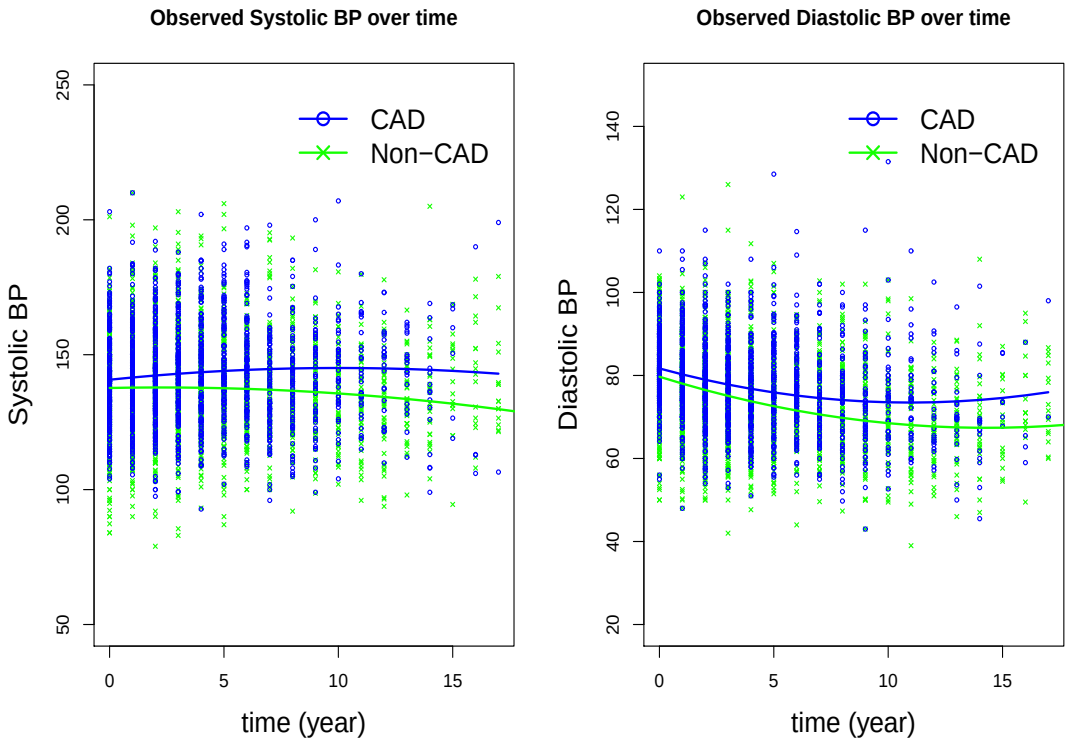

Figure 3.1: Observed annualized longitudinal systolic and diastolic BP measures over time and fitted population mean curves for the CAD and non-CAD group.

CAD and non-CAD groups were negligible. Thus analyses relying on baseline BP measures may not be able to detect any relationship between BP measures and risk of CAD.

3.4 Joint Models

In this section, we introduce joint models for multiple longitudinal processes and a time-to-event outcome by defining the notations and formulation of the longitudinal and survival models. Specifically, we consider multivariate mixed-effects models for the multiple longitudinal outcomes and a Cox model for the time-to-event outcome with predicted functions

of the longitudinal measures as time-dependent covariates. We then derive the likelihood function of the joint models.

3.4.1 Longitudinal Models

Let $y_l(t_{ij})$ denote the observed measurement of the l-th longitudinal outcome for subject i at time points t_{ij}, where $i = 1, ..., n$, $j = 1, ..., n_i$, $l = 1, ..., L$. The corresponding longitudinal trajectory is modeled using the following model

$$\begin{aligned} \mathbf{y}_l(\mathbf{t}_i) &= \mathbf{y}_l^*(\mathbf{t}_i) + \boldsymbol{\epsilon}_{il}, \\ &= \mathbf{X}_l^T(\mathbf{t}_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(\mathbf{t}_i)\mathbf{b}_{il} + \boldsymbol{\epsilon}_{il} \end{aligned} \tag{3.1}$$

where $\mathbf{y}_l^*(\mathbf{t}_i) = (y_l(t_{i1}), y_l(t_{i2}), ..., y_l(t_{in_i}))^T$ is the corresponding true underlying longitudinal measures of the l-th biomarker for the i-th subject; $\mathbf{X}_l^T(\mathbf{t}_i)$ is the design matrix of fixed effects, including time effects and baseline covariates; $\boldsymbol{\beta}_l$ is the corresponding vector of the fixed effects; $\mathbf{Z}_l^T(\mathbf{t}_i)$ is the design matrix for the random effects, $\mathbf{b}_{il}$, distributed as $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, ..., \mathbf{b}_{iL})^T \sim N(\mathbf{0}, \mathbf{D})$; $\boldsymbol{\epsilon}_{il}$ is the corresponding measurement error term such that $\boldsymbol{\epsilon}_{il} \sim^{iid} N(0, \sigma_l^2 I_{n_i})$. It is worth noting that the correlations among the multiple longitudinal processes and the within-subject correlation for each longitudinal biomarker are represented in the variance-covariance matrix of random effects $\mathbf{D}$. We assume that the measurement errors of different longitudinal outcomes are independent of each other, and they are also independent of the random effects $\mathbf{b}_i$.

3.4.2 The Survival Model

Let T_i^* and C_i be the true event time and censoring time respectively for subject i. We define the observed event time $T_i = \min(T_i^*, C_i)$ and the event indicator $\delta_i = I(T_i^* \leq C_i)$. Assuming that the hazard function depends on some functions of the true longitudinal

measures $\mathcal{F}(\mathbf{y}_{il}^*(t))$ and baseline covariates $\mathbf{w_i}$, the hazard function can be written as

$$h_i(t) \quad = \quad h_0(t) \exp \left\{ \boldsymbol{\gamma}^T \mathbf{w_i} + \sum_{l=1}^{L} \alpha_l \mathcal{F}(\mathbf{y}_{il}^*(t)) \right\}, \tag{3.2}$$

where $h_0(t)$ denotes the baseline hazard function, and α_l and $\boldsymbol{\gamma}$ are coefficients for the function of lth biomarker and baseline risk factors. The baseline hazard function can be a parametric function or a flexible piecewise constant function. In this work, α_l, $l = 1, 2, ..., L$, are of primary interest. The correlation between the multiple longitudinal biomarkers and the time-to-event outcome is induced by the shared random effects through $y_{il}^*(t)$ or b_{il} in the longitudinal and survival models.

The function $\mathcal{F}(\cdot)$ can be chosen as different functional forms depending on the interest of the study. For example, if the focus is the association between longitudinal values and event risk, $\mathcal{F}(\cdot)$ can be an identity function; if the change in the longitudinal measures is of interest, $\mathcal{F}(\cdot)$ can be chosen as derivative function with respect to time t; for studies interested in the cumulative history of the longitudinal measures over time and event risk, $\mathcal{F}(\cdot)$ can be an integration function of $\mathbf{y}_{il}^*(t)$ over time t. Depending on the choices, random effects b_i may affect the hazard function in a non-linear fashion. This is in contrast with the frailty type of joint models where the random effects are linear in the exponential term of the hazard.

3.4.3 Joint Likelihood Function

Under the conditional independence assumption between $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ the kernel of the joint likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} p(T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}|\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \int p(T_i, \delta_i|\boldsymbol{\theta}, \mathbf{b}_i) p(\mathbf{y}_{i1}|\boldsymbol{\theta}, \mathbf{b}_i) \cdots p(\mathbf{y}_{iL}|\boldsymbol{\theta}, \mathbf{b}_i) p(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i, \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\theta}_{h_0}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{D})^T$ is the vector containing all parameters in the models. Under the assumed models (3.1) and (3.2) for the longitudinal and survival outcomes, there are three components:

$$\begin{aligned} p(T_i, \delta_i|\boldsymbol{\theta}, b_i) = &\left\{ h_0(T_i) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il} \Big) \right) \right\}^{\delta_i} \\ &\exp\left\{ - \int_0^{T_i} h_0(u) \exp\left(\mathbf{w}_i^T \gamma + \sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_l^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(u)\mathbf{b}_{il} \Big) \right) du \right\}, \end{aligned}$$

$$\begin{aligned} &p(\mathbf{y}_{i1}|\boldsymbol{\theta}, \mathbf{b}_i) \cdots p(\mathbf{y}_{iL}|\boldsymbol{\theta}, \mathbf{b}_i) \\ &= \prod_{l=1}^{L} \left(\frac{1}{\sqrt{2\pi\sigma_l^2}} \right)^{n_i} \exp\left\{ -\frac{1}{2\sigma_l^2} \sum_{j=1}^{n_i} \Big(\mathbf{y}_l(t_{ij}) - \Big(\mathbf{X}_l^T(t_{ij})\boldsymbol{\beta}_l + \mathbf{Z}_l^T(t_{ij})\mathbf{b}_{il} \Big) \Big)^2 \right\}, \end{aligned}$$

and

$$p(\mathbf{b}_i|\boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{k/2} |\mathbf{D}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i \right),$$

where k is the dimension of the $\mathbf{D}$ matrix.

3.5 Estimation Method

In this section, we present the maximum-likelihood method based on the EM algorithm for parameter estimation and inference.

3.5.1 Implementing the EM Algorithm

The complete log likelihood function given the random effects $\mathbf{b}_i$ is:

$$\begin{aligned}\log L_C &= \sum_{i=1}^{n}\Big\{\log p(T_i,\delta_i|\mathbf{b}_i;\boldsymbol{\theta})+\log p(\mathbf{y}_{i1}|\mathbf{b}_i;\boldsymbol{\theta})+...+\log p(\mathbf{y}_{iL}|\mathbf{b}_i;\boldsymbol{\theta})\\ &\quad+\log p(\mathbf{b}_i;\boldsymbol{\theta})\Big\}.\end{aligned}$$

In the E-step the expected complete log-likelihood function given the conditional distribution of random effects is

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^m)=\sum_{i=1}^{n}\int\Big\{&\log p(T_i,\delta_i|\mathbf{b}_i;\boldsymbol{\theta})+\log p(\mathbf{y}_{i1}|\mathbf{b}_i;\boldsymbol{\theta})+...+\log p(\mathbf{y}_{iL}|\mathbf{b}_i;\boldsymbol{\theta})\\ &+\log p(\mathbf{b}_i;\boldsymbol{\theta})\Big\}p(\mathbf{b}_i|T_i,\delta_i,\mathbf{y}_{i1},...,\mathbf{y}_{iL};\boldsymbol{\theta}^m)d\mathbf{b}_i.\end{aligned}$$

For the M-step, close forms are available for the variance of residuals of each longitudinal model and variance-covariance matrix of the random effects, whereas the fixed effects for each longitudinal model and parameters in the survival model have to be estimated numerically. The key steps are:

1. Estimation of the variance of residuals of each longitudinal model by

$$\begin{aligned}\hat{\sigma_l}^2=\frac{1}{\sum_{i=1}^{n}n_i}\sum_{i=1}^{n}&\Big(\mathbf{y}_{il}-\mathbf{X}_{il}^T\boldsymbol{\beta}_l^m\Big)^T\Big(\mathbf{y}_{il}-\mathbf{X}_{il}^T\boldsymbol{\beta}_l^m-2\mathbf{Z}_{il}^T\mathrm{E}(\mathbf{b}_i|T_i,\delta_i,\mathbf{y}_{i1},...,\mathbf{y}_{iL};\boldsymbol{\theta}^m)\Big)\\ &+\mathrm{Tr}\Big(\mathbf{Z}_{il}^T\mathbf{Z}_{il}\mathrm{Var}(\mathbf{b}_i|T_i,\delta_i,\mathbf{y}_{i1},...,\mathbf{y}_{iL};\boldsymbol{\theta}^m)\Big)\\ &+\mathrm{E}(\mathbf{b}_i|T_i,\delta_i,\mathbf{y}_{i1},...,\mathbf{y}_{iL};\boldsymbol{\theta}^m)^T\mathbf{Z}_{il}^T\mathbf{Z}_{il}\mathrm{E}(\mathbf{b}_i|T_i,\delta_i,\mathbf{y}_{i1},...,\mathbf{y}_{iL};\boldsymbol{\theta}^m)\end{aligned}$$

   where $l=1,...,L$, Tr represents the trace function of a matrix, and E denotes the expectation function.

2. Estimation of variance-covariance matrix of random effects by

$$\hat{\mathbf{D}} = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \mathrm{Var}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) + \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m)^T$$

3. Since the fixed effect coefficients of each longitudinal model, $\boldsymbol{\beta}_l$, are involved in both the longitudinal and survival models, there is no closed form solution. The one-step Newton-Raphson algorithm can be implemented to update $\boldsymbol{\beta}_l$:

$$\hat{\boldsymbol{\beta}}_l^{m+1} = \hat{\boldsymbol{\beta}}_l^{m} - \left(\partial S(\hat{\boldsymbol{\beta}}_l^{m}) / \partial \boldsymbol{\beta}_l \right)^{-1} S(\hat{\boldsymbol{\beta}}_l^{m}), l = 1, ..., L,$$

where the score functions are

$$\begin{aligned} S(\boldsymbol{\beta}_l) = \sum_{i=1}^{n} & \frac{1}{\sigma_l^2} \mathbf{X}_{il}^T \left(\mathbf{y}_{il} - \mathbf{X}_{il}^T \boldsymbol{\beta}_l - \mathbf{Z}_{il}^T \mathrm{E}(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) \right) \\ & + \delta_i \alpha_l \frac{\partial \mathcal{F}\left(\mathbf{X}_l^T(T_i) \boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i) \mathbf{b}_{il} \right)}{\partial \boldsymbol{\beta}_l} \\ & - \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \frac{\partial \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\left(\mathbf{X}_{il}^T(u) \boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u) \mathbf{b}_i \right) \right)}{\partial \boldsymbol{\beta}_l} \\ & p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i, \end{aligned}$$

for $l = 1, ..., L$. The derivatives $\partial S(\hat{\boldsymbol{\beta}}_l^{m}) / \partial \boldsymbol{\beta}_l$ can be calculated by numerical approximations.

4. Parameters in the survival model can be similarly updated using the Newton-Raphson algorithm. Score equations used in the Newton-Raphson algorithm are:

$$\begin{aligned} S(\gamma) = \sum_{i=1}^{n} \mathbf{w}_i \Bigg\{ & \delta_i - \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\left(\mathbf{X}_{il}^T(u) \boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u) \mathbf{b}_i \right) \right) \\ & p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i \Bigg\}. \end{aligned}$$

$$S(\alpha_l) = \sum_{i=1}^{n} \delta_i \int \mathcal{F}\Big(\mathbf{X}_l^T(T_i)\boldsymbol{\beta}_l + \mathbf{Z}_l^T(T_i)\mathbf{b}_{il}\Big) d\mathbf{b}_i$$
$$- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} h_0(u) \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i\Big)$$
$$\exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i\Big)\right) p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i,$$

where $l = 1, ..., L$, and

$$S(\boldsymbol{\theta}_{h_0}) = \sum_{i=1}^{n} \delta_i \frac{1}{h_0(T_i; \boldsymbol{\theta}_{h_0})} \frac{\partial h_0(T_i; \boldsymbol{\theta}_{h_0})}{\partial \boldsymbol{\theta}_{h_0}^T}$$
$$- \exp(\gamma^T \mathbf{w}_i) \int \int_0^{T_i} \frac{\partial h_0(T_i; \boldsymbol{\theta}_{h_0})}{\partial \boldsymbol{\theta}_{h_0}^T} \exp\left(\sum_{l=1}^{L} \alpha_l \mathcal{F}\Big(\mathbf{X}_{il}^T(u)\boldsymbol{\beta}_l + \mathbf{Z}_{il}^T(u)\mathbf{b}_i\Big)\right)$$
$$p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; \boldsymbol{\theta}^m) du d\mathbf{b}_i.$$

For computation of the expected likelihood function, a pseudo-adoptive Gaussian-Hermit quadrature rule (Rizopoulos, 2012a) can be used to approximate the integrals. The E-step and M-step iterate until a pre-specified convergence criterion is met.

3.5.2 Inferences and Goodness-of-fit

After convergence of the EM algorithm, standard errors of estimated parameters can be calculated from the observed inverse Hessian matrix, which can be derived from the score functions using numeric approximations Press et al. (2007). Using the properties of consistency and asymptotic normality of maximum-likelihood estimates, a 95% confidence interval (CI) of a parameter θ can be derived as:

$$\hat{\theta} \pm 1.96 \times s.e.(\hat{\theta}).$$

For hypothesis tests involving multiple parameters,

$$\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \;\; vs. \;\; \mathrm{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,$$

the Wald test or Score test can be used. The Wald test statistic is given by

$$\mathrm{W} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0)^T I(\hat{\boldsymbol{\theta}_0})(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0),$$

where $I(\hat{\boldsymbol{\theta}_0})$ denotes the observed information matrix under the null hypothesis. The Score test statistic is given by

$$\mathrm{U} = S^T(\hat{\boldsymbol{\theta}}_0)\{I(\hat{\boldsymbol{\theta}}_0)\}^{-1} S(\hat{\boldsymbol{\theta}}_0),$$

where $S(\hat{\boldsymbol{\theta}}_0)$ and $I(\hat{\boldsymbol{\theta}}_0)$ are the observed score function and information matrix under the null hypothesis.

The Likelihood Ratio Test (LRT) is a commonly used statistic to compare nested models, given by

$$\mathrm{LRT} = -2\{l(\hat{\boldsymbol{\theta}_0}) - l(\hat{\boldsymbol{\theta}})\},$$

where $l()$ is the log joint likelihood function, $\hat{\boldsymbol{\theta}_0}$ and $\hat{\boldsymbol{\theta}}$ are the corresponding maximum-likelihood estimates under the null and alternative hypotheses respectively. For non-nested models, the Akaike's Information Criterion (AIC) Akaike (1987) and Bayesian Information Criterion (BIC) Schwarz (1978) are often used for model selection:

$$\mathrm{AIC} = -2l(\hat{\boldsymbol{\theta}}) + 2k,$$

$$\mathrm{BIC} = -2l(\hat{\boldsymbol{\theta}}) + k\log(n),$$

where k denotes the number of parameters in the model and n is the number of subjects in the data.

3.6 Simulation Study

We performed Monte Carlo (MC) simulations to assess the performance of the proposed method and compare with the two-stage method. We simulated data from joint models with two correlated normally distributed longitudinal variables and a time-to-event variable. For each MC data set, longitudinal data were simulated for 500 subjects each with 10 equally spaced bivariate longitudinal measures over a 5-year period. We considered similar fixed and random model structures for the two longitudinal outcomes, where the fixed effects included time effect, and one binary baseline covariate, and the random effects included random intercept and random slope. The correlation between the two longitudinal outcomes is represented by the correlation between the two random intercepts. The longitudinal models are:

$$y_1(t_{ij}) = y_1^*(t_{ij}) + \epsilon_1(t_{ij}) = \beta_{01} + \beta_{11}t_{ij} + \beta_{21}w_i + b_{01i} + b_{11i}t_i + \epsilon_1(t_{ij}),$$

$$y_2(t_{ij}) = y_2^*(t_{ij}) + \epsilon_2(t_{ij}) = \beta_{02} + \beta_{12}t_{ij} + \beta_{22}w_i + b_{02i} + b_{12i}t_i + \epsilon_2(t_{ij}),$$

where $\epsilon_1(t_{ij}) \sim N(0, \sigma_1^2)$, $\epsilon_2(t_{ij}) \sim N(0, \sigma_2^2)$ and

$$(b_{01i}, b_{11i}, b_{02i}, b_{12i})^T \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{01}^2 & 0 & \rho\sigma_{01}\sigma_{02} & 0 \\ 0 & \sigma_{11}^2 & 0 & 0 \\ \rho\sigma_{01}\sigma_{02} & 0 & \sigma_{02}^2 & 0 \\ 0 & 0 & 0 & \sigma_{12}^2 \end{pmatrix}\right).$$

The time-to-event endpoint was simulated from a Cox model with a Weibull baseline hazard function, $h_0(t) = abt^{b-1}$, where a and b are the shape and scale parameters respectively. The Cox model is assumed to depend on the current values of the two longitudinal outcomes and can be expressed as

$$h(t) = abt^{b-1} \exp(\alpha_1 y_1^*(t) + \alpha_2 y_2^*(t)).$$

To simulate the event times, we first simulated a survival probability, s_i, from Uniform(0,1) for each subject and then solved for T_i^* using two R functions R Development Core Team (2007), *integrate*() and *uniroot*(), from the following equation:

$$s_i - \exp\left\{ - \int_0^{T_i^*} abu^{b-1} \exp(\alpha_1 y_1^*(u) + \alpha_2 y_2^*(u)) du \right\} = 0.$$

Censoring times were independently simulated from another uniform distribution. Overall, the censoring percentage is about 30%. Because of censoring, there were about 6 repeated bivariate measurements per subject. In the actual model fitting, we used a more flexible piecewise constant baseline hazard function instead of the parametric Weibull baseline risk function.

Several simulation scenarios were considered, varying in variances of residual errors, variance-covariance matrix of the random effects, and the correlation between the two longitudinal outcomes. Specifically, four scenarios were considered. The true parameter values used in the simulations were summarized in Table 3.1 and Table 3.2.

Simulation results are presented in Tables 3.3, 3.4, 3.5 and 3.6. We reported relative biases, empirical standard errors, model-based standard errors, and coverage probabilities of the 95% CIs based on 500 MC data sets. For the variance-covariance matrix of the

Table 3.1: True parameter values for the four scenarios of simulation studies.

Scenario 1		Scenario 2		Scenario 3		Scenario 4	
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
σ_1	0.1	σ_1	0.5	σ_1	0.1	σ_1	0.5
σ_2	0.1	σ_2	0.5	σ_2	0.1	σ_2	0.5
σ_{01}	0.1	σ_{01}	0.5	σ_{01}	0.1	σ_{01}	0.5
σ_{11}	0.04	σ_{11}	0.2	σ_{11}	0.04	σ_{11}	0.2
σ_{02}	0.1	σ_{02}	0.5	σ_{02}	0.1	σ_{02}	0.5
σ_{12}	0.04	σ_{12}	0.2	σ_{12}	0.04	σ_{12}	0.2
ρ	0.2	ρ	0.2	ρ	0.5	ρ	0.5

Table 3.2: True parameter values for the two longitudinal models and proportional hazard function of simulation studies.

Outcome 1		Outcome 2		Time-to-event outcome	
Parameter	Value	Parameter	Value	Parameter	Value
β_{01}	0.2	β_{02}	1.0	a	0.005
β_{11}	0.5	β_{12}	0.2	b	1.1
β_{21}	0.2	β_{22}	0.5	α_1	1.0
				α_2	1.5

random effects, $\mathbf{D}$, we reported the results of its cholesky decomposition, where

$$chol(\mathbf{D}) = \begin{pmatrix} D_{11} & 0 & D_{13} & 0 \\ 0 & D_{22} & 0 & 0 \\ 0 & 0 & D_{33} & 0 \\ 0 & 0 & 0 & D_{44} \end{pmatrix}.$$

For scenarios 1 and 3 (small variances), both the two-stage approach and EM algorithm generally performed well: estimated parameters have low relative bias (defined as $(\boldsymbol{\theta}_{\text{true}} - \hat{\boldsymbol{\theta}})/\boldsymbol{\theta}_{\text{true}}$), and coverage probability rates of 95% CIs are around the nominal level. However, in scenarios 2 and 4 (large variances), the two-stage approach leads to poor results whereas

the EM algorithm maintains good performance with small relative bias and good coverage probability rates of 95% CIs. In particular, the two-stage approach tends to underestimate the two association parameters, α_1 and α_2 . In addition, it is observed the correlations between the two longitudinal outcomes have little impact on the parameter estimation.

We also conducted a simulation study to compare the performance of the EM algorithm with small and large censoring percentages. We simulated 500 MC samples with 60% censoring percentage using the same parameter setup as in scenario 4. Results are presented in Table 3.7, showing that the EM algorithm performs satisfactorily even with 60% censoring data.

Table 3.3: Simulation results for comparing Joint model approach and Two-stage method under scenario 1 with censoring percentage(30%).

		Joint Model				Two-stage			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	0.000	0.008	0.008	95.6%	-0.004	0.008	0.008	94.6%
β_{11}	0.50	-0.000	0.003	0.003	96%	0.002	0.003	0.003	93.2%
β_{21}	0.20	0.001	0.011	0.011	94%	0.003	0.011	0.011	94.8%
$\log(\sigma_1)$	-2.30	-0.000	0.017	0.017	95%	0.000	0.017	0.016	95.6%
β_{02}	1.00	0.000	0.008	0.007	94.4%	-0.001	0.008	0.007	94.8%
β_{12}	0.20	0.000	0.003	0.003	94.4%	0.009	0.003	0.003	88.8%
β_{22}	0.50	-0.001	0.011	0.010	96.2%	-0.000	0.011	0.010	96.4%
$\log(\sigma_2)$	-2.30	0.000	0.017	0.017	96.6%	0.000	0.017	0.016	95.6%
α_1	1.00	-0.103	0.198	0.197	91.8%	-0.010	0.375	0.386	94.2%
α_2	1.50	-0.002	0.218	0.217	95%	-0.003	0.233	0.230	95%
D_{11}	-2.30	-0.002	0.050	0.050	95.6%	-0.001	0.044	0.045	94.4%
D_{22}	-3.22	-0.005	0.079	0.079	95%	-0.003	0.070	0.069	94.6%
D_{13}	0.02	0.015	0.007	0.007	94.8%	0.026	0.005	0.006	89%
D_{33}	-2.32	-0.005	0.052	0.055	94.6%	-0.002	0.044	0.048	92.8%
D_{44}	-3.22	-0.008	0.081	0.085	97.4%	-0.003	0.070	0.071	96.2%

Table 3.4: Simulation results for comparing the EM algorithm and the two-stage approach under scenario 2 with censoring percentage(30%).

		EM				Two-stage			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	-0.041	0.038	0.040	91%	-0.094	0.037	0.040	88.2%
β_{11}	0.50	-0.002	0.017	0.016	93.2%	0.045	0.015	0.015	78.2%
β_{21}	0.20	0.035	0.052	0.055	94.2%	0.072	0.052	0.054	93.8%
$\log(\sigma_1)$	-0.69	0.012	0.017	0.018	93.6%	0.014	0.017	0.018	93%
β_{02}	1.00	-0.005	0.038	0.039	90.2%	-0.022	0.037	0.040	82.2%
β_{12}	0.20	-0.021	0.016	0.017	92.6%	0.160	0.014	0.015	43.8%
β_{22}	0.50	0.032	0.051	0.052	94%	0.054	0.051	0.052	92%
$\log(\sigma_2)$	-0.69	0.013	0.017	0.017	94.2%	0.014	0.017	0.018	93%
α_1	1.00	-0.014	0.103	0.110	92.6%	0.137	0.090	0.097	67.4%
α_2	1.50	0.010	0.126	0.124	94.6%	0.116	0.101	0.101	61.6%
D_{11}	-0.69	-0.001	0.049	0.052	94.2%	-0.012	0.045	0.046	93.2%
D_{22}	-1.61	-0.020	0.079	0.084	90.6%	-0.025	0.072	0.076	89%
D_{13}	0.10	0.037	0.034	0.034	95.8%	0.234	0.024	0.030	78.4%
D_{33}	-0.71	-0.020	0.052	0.052	93.2%	-0.033	0.045	0.047	89%
D_{44}	-1.61	-0.060	0.083	0.080	75.6%	-0.077	0.078	0.077	57.4%

Table 3.5: Simulation results for comparing the EM algorithm and the two-stage approach under scenario 3 with censoring percentage(30%).

		EM				Two-stage			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	0.000	0.008	0.008	95.4%	-0.004	0.008	0.007	95.2%
β_{11}	0.50	-0.000	0.003	0.003	94.2%	0.003	0.003	0.003	92.6%
β_{21}	0.20	-0.000	0.011	0.010	96.8%	0.002	0.011	0.010	97%
$\log(\sigma_1)$	-2.30	0.000	0.017	0.017	94.8%	0.000	0.017	0.017	94.8%
β_{02}	1.00	0.001	0.008	0.008	95%	-0.001	0.008	0.008	95.4%
β_{12}	0.20	0.001	0.003	0.003	94.4%	0.010	0.003	0.003	91.6%
β_{22}	0.50	-0.001	0.011	0.010	95%	0.000	0.011	0.010	95.2%
$\log(\sigma_2)$	-2.30	0.000	0.017	0.016	96.6%	0.000	0.017	0.017	94.8%
α_1	1.00	-0.104	0.201	0.198	92.6%	-0.000	0.393	0.378	95.4%
α_2	1.50	-0.004	0.221	0.223	94.2%	-0.007	0.243	0.242	94.6%
D_{11}	-2.30	-0.003	0.050	0.049	94.8%	-0.002	0.044	0.043	96.6%
D_{22}	-3.22	-0.004	0.079	0.082	95.6%	-0.001	0.069	0.070	95.4%
D_{13}	0.05	-0.003	0.007	0.007	94.4%	-0.001	0.006	0.006	91.4%
D_{33}	-2.45	-0.007	0.064	0.065	95.4%	-0.002	0.044	0.054	89%
D_{44}	-3.22	-0.009	0.082	0.083	96%	-0.003	0.069	0.069	94.4%

Table 3.6: Simulation results for comparing the EM algorithm and the two-stage approach under scenario 4 with censoring percentage(30%).

		EM				Two-stage			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	-0.062	0.038	0.037	93.4%	-0.109	0.038	0.037	91.4%
β_{11}	0.50	-0.014	0.017	0.017	94.6%	0.034	0.015	0.015	78.8%
β_{21}	0.20	0.041	0.052	0.055	94.6%	0.077	0.052	0.054	93.8%
$\log(\sigma_1)$	-0.69	0.003	0.017	0.018	95%	0.005	0.017	0.018	93.8%
β_{02}	1.00	-0.016	0.038	0.041	91.8%	-0.032	0.037	0.040	83%
β_{12}	0.20	-0.041	0.017	0.016	93%	0.135	0.015	0.014	55.6%
β_{22}	0.50	0.019	0.052	0.054	92.4%	0.040	0.051	0.054	91.6%
$\log(\sigma_2)$	-0.69	0.001	0.017	0.017	96.2%	0.005	0.017	0.018	93.8%
α_1	1.00	-0.035	0.106	0.103	95.6%	0.122	0.093	0.090	71.6%
α_2	1.50	0.010	0.127	0.123	95.6%	0.114	0.103	0.101	59.2%
D_{11}	-0.69	-0.016	0.050	0.054	94.2%	-0.036	0.046	0.049	92%
D_{22}	-1.61	-0.033	0.079	0.081	91%	-0.037	0.072	0.074	86.6%
D_{13}	0.25	0.014	0.033	0.033	95%	0.109	0.026	0.029	78%
D_{33}	-0.84	-0.027	0.064	0.062	96.6%	-0.023	0.046	0.050	91%
D_{44}	-1.61	-0.070	0.085	0.083	80.6%	-0.082	0.077	0.076	65%

Table 3.7: Simulation results for comparing the EM algorithm under scenarios of small and large censoring percentage: 30% VS 60%.

		30%				60%			
Parameter	True	Relative Bias	Model SE	Emp. SE	95%CI Cov.	Relative Bias	Model SE	Emp. SE	95%CI Cov.
β_{01}	0.20	-0.062	0.038	0.037	93.4%	-0.083	0.039	0.042	91%
β_{11}	0.50	-0.014	0.017	0.017	94.6%	-0.016	0.020	0.021	91.6%
β_{21}	0.20	0.041	0.052	0.055	94.6%	0.057	0.053	0.055	93.8%
$\log(\sigma_1)$	-0.69	0.003	0.017	0.018	95%	0.002	0.021	0.021	93.8%
β_{02}	1.00	-0.016	0.038	0.041	91.8%	-0.019	0.039	0.039	90.6%
β_{12}	0.20	-0.041	0.017	0.016	93%	-0.044	0.020	0.019	93.2%
β_{22}	0.50	0.019	0.052	0.054	92.4%	0.021	0.053	0.053	93.8%
$\log(\sigma_2)$	-0.69	0.001	0.017	0.017	96.2%	0.001	0.020	0.021	95.2%
α_1	1.00	-0.035	0.106	0.103	95.6%	-0.035	0.155	0.158	95.4%
α_2	1.50	0.010	0.127	0.123	95.6%	-0.012	0.175	0.169	96%
D_{11}	-0.69	-0.016	0.050	0.054	94.2%	-0.016	0.053	0.053	94%
D_{22}	-1.61	-0.033	0.079	0.081	91%	-0.046	0.112	0.223	96.8%
D_{13}	0.25	0.014	0.033	0.033	95%	0.024	0.035	0.035	94.4%
D_{33}	-0.84	-0.027	0.064	0.062	96.6%	-0.028	0.072	0.068	97.6%
D_{44}	-1.61	-0.070	0.085	0.083	80.6%	-0.082	0.122	0.175	91.4%

3.7 Data Application

We applied joint models to the aforementioned primary care patient cohort data using 797 male patients. Our focus was to determine whether and how longitudinal BP measures were associated with the time to CAD. For convenience we centered patients' baseline age at 60 years. We fitted four different sets of joint models using the proposed EM algorithm. The best set of models were determined using the AIC. The 4 sets of joint models are as follows.

Joint models 1 consider the following models

$$\begin{aligned} y_l(t_{ij}) &= y_l^*(t_{ij}) + \epsilon_{ijl} \\ &= \beta_{0l} + \beta_{1l}t_{ij} + \beta_{2l}\text{age}_i + \beta_{3l}\text{race}_i + b_{0li} + b_{1li}t_{ij} + \epsilon_{ijl}, l = 1, 2 \end{aligned} \tag{3.3}$$

for the observed longitudinal systolic and diastolic BP measures respectively. In the above longitudinal model, age and race, two well-known factors correlated with systolic and diastolic BP are included as fixed effects. The random effect vector $\mathbf{b}_i = (b_{01i}, b_{11i}, b_{02i}, b_{12i})^T$ is normally distributed with mean zero and an unstructured variance-covariance matrix of $\mathbf{D}$; ϵ_{ij1} and ϵ_{ij2} are independently distributed as $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively. The hazard function satisfies

$$h(t) = h_0(t) \exp\left\{\gamma_1\text{age}_i + \gamma_2\text{smoke}_i + \gamma_3\text{race}_i + \gamma_4\text{diabetes}_i + \alpha_1 y_{i1}^*(t) + \alpha_2 y_{i2}^*(t)\right\} \tag{3.4}$$

where the piecewise constant function $h_0(t)$ consists of 7 equally spaced intervals with 6 interior knots based on percentiles of the observed event time points. The above hazard function depends on the current values of systolic and diastolic BP measures and some common risk factors of CAD, age, smoking history, race and diabetes status.

Joint models 2 assume the same longitudinal models (3.3) as in the joint models 1, but use the slopes of systolic and diastolic BP measures in the hazard function instead,

$$h(t) = h_0(t) \exp\left\{\gamma_1 \text{age}_i + \gamma_2 \text{smoke}_i + \gamma_3 \text{race}_i + \gamma_4 \text{diabetes}_i + \alpha_1 y_{i1}^{*'}(t) + \alpha_2 y_{i2}^{*'}(t)\right\} \quad (3.5)$$

Joint models 3 assume the same hazard model (3.4) in the joint model 1, but include a quadratic fixed time effect for both systolic and diastolic BP measures,

$$\begin{aligned} y_l(t_{ij}) &= y_l^*(t_{ij}) + \epsilon_{ijl} \\ &= \beta_{0l} + \beta_{1l} t_{ij} + \underline{\beta_{2l} t_{ij}^2} + \beta_{3l} \text{age}_i + \beta_{4l} \text{race}_i + b_{0li} + b_{1li} t_{ij} + \epsilon_{ijl}, l = 1, 2. \end{aligned} \quad (3.6)$$

Joint models 4 assumes (3.5) and (3.6).

In the implementation of the EM algorithm, we used 3 pseudo-adaptive Gaussian-Hermite quadrature points for numerical integration over the random effects and 7 Gaussian-Kronrod quadrature points for the integration in the survival function. Estimated parameters for the four sets of joint models are presented in Tables 3.8, 3.9, 3.10, and 3.11 respectively. Models were compared according to AIC: smaller AIC indicates better model fit. Among the 4 joint models considered, Joint models 3 was the best fitting (AIC=65786) followed by Joint models 4 (AIC=65798), Joint models 2 (AIC=65885) and Joint models 1 (AIC=65898). Here we focus on Joint models 3 for inference and interpretation.

It can be seen that systolic BP measures are significantly associated with the risk of developing CAD. Each 10 unit increase of systolic BP is associated with 1.23-fold increase (95% CI:[1.05, 1.5]) in patient's risk of developing CAD. In addition we observe that diastolic BP measures are not significantly associated with the risk of developing CAD once systolic BP measures were adjusted in the model. The fitted model also identified several other risk factors for CAD, i.e. participants with older age, being Caucasian and smokers

have higher risk of CAD. The fitted longitudinal quadratic growth models suggested that there is a quadratic increasing-then-decreasing trend for systolic BP measures, whereas a decreasing-then-increasing quadratic trend for diastolic BP measures was seen. In Figure3.2 we plotted subject-specific fitted curves under fitted Joint model 3 for 4 CAD and 4 non-CAD participants, randomly selected from the study population. It can be seen that the quadratic longitudinal models fit the data relatively well.

As a comparison, we also fitted two separate single longitudinal measure joint models, one using systolic BP only and the other diastolic BP only, while adjusting for the same covariates as in (3.6) above. The two separate joint models showed that both systolic and diastolic BP were significantly associated with CAD risk. Our joint models 3 takes the correlation between the two BP measures into consideration and our results indicate that systolic BP had higher impact on the risk of CAD than diastolic BP in this elderly population.

We also analyzed the data using alternative methods including the two-stage approach, Cox model with baseline BP measures, and Cox model with time-dependent BP measures. In the two Cox models we adjusted for the same baseline risk factors as in (3.4). Figure 3.3 plots the estimated parameter, $\hat{\alpha}_1$, for systolic BP and their corresponding 95% CIs from 4 different methods: the EM algorithm (Joint models 3), the two-stage method (Joint models 3), the Cox model with time-dependent covariates and the Cox model with baseline BP measures. It can be seen that joint models using the EM algorithm has the largest parameter estimate among all the methods. The under estimation of the two-stage method was expected given the results from the simulation study. Nevertheless, both joint modeling approaches (EM algorithm and two-stage approach) point to stronger associations between systolic BP and CAD than the two Cox models.

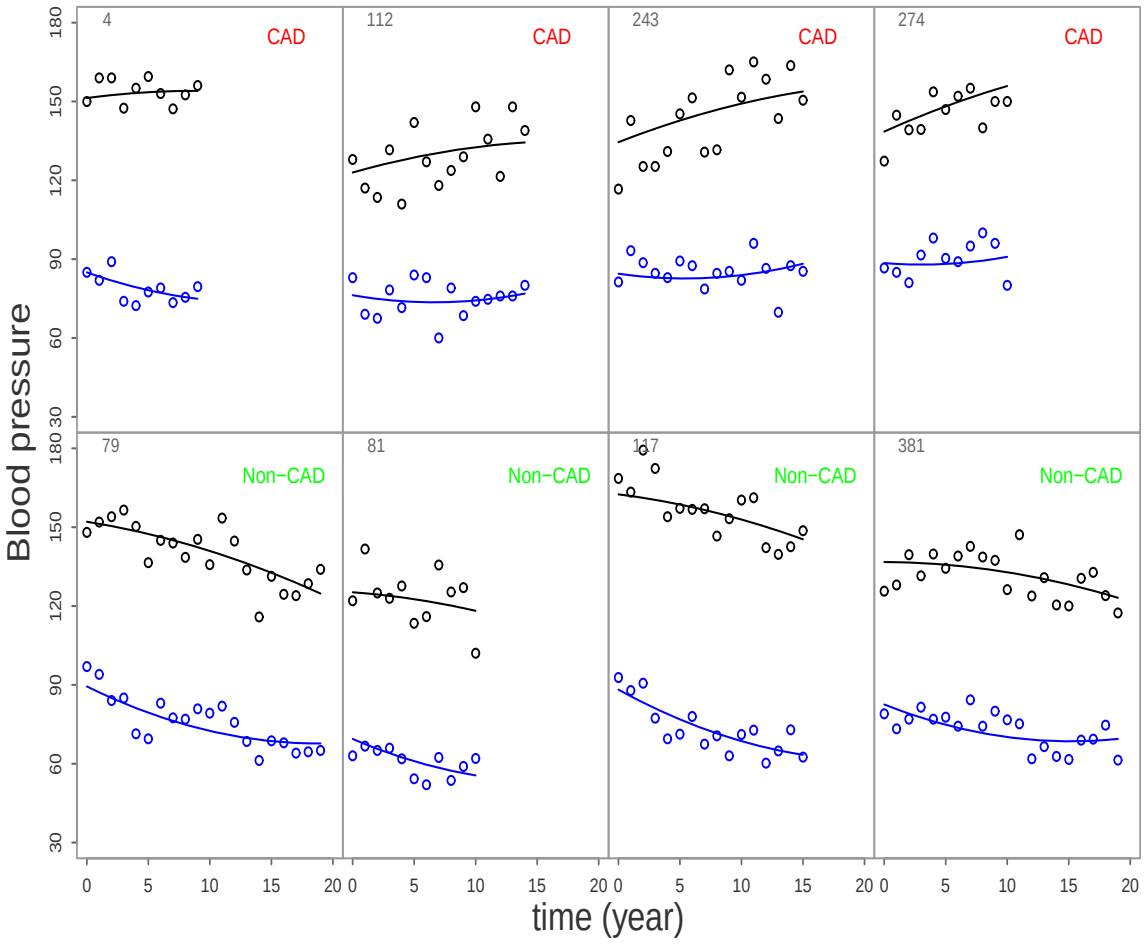

Figure 3.2: Fitted subject-specific longitudinal BP curves for randomly selected 4 CAD and 4 non-CAD subjects based on fitted Joint models 3. The black dots and black solid curves represent the observed systolic BP overtime and fitted subject-specific curves respectively. The blue dots and blue solid curves represent the observed diastolic BP overtime and fitted subject-specific curves respectively.

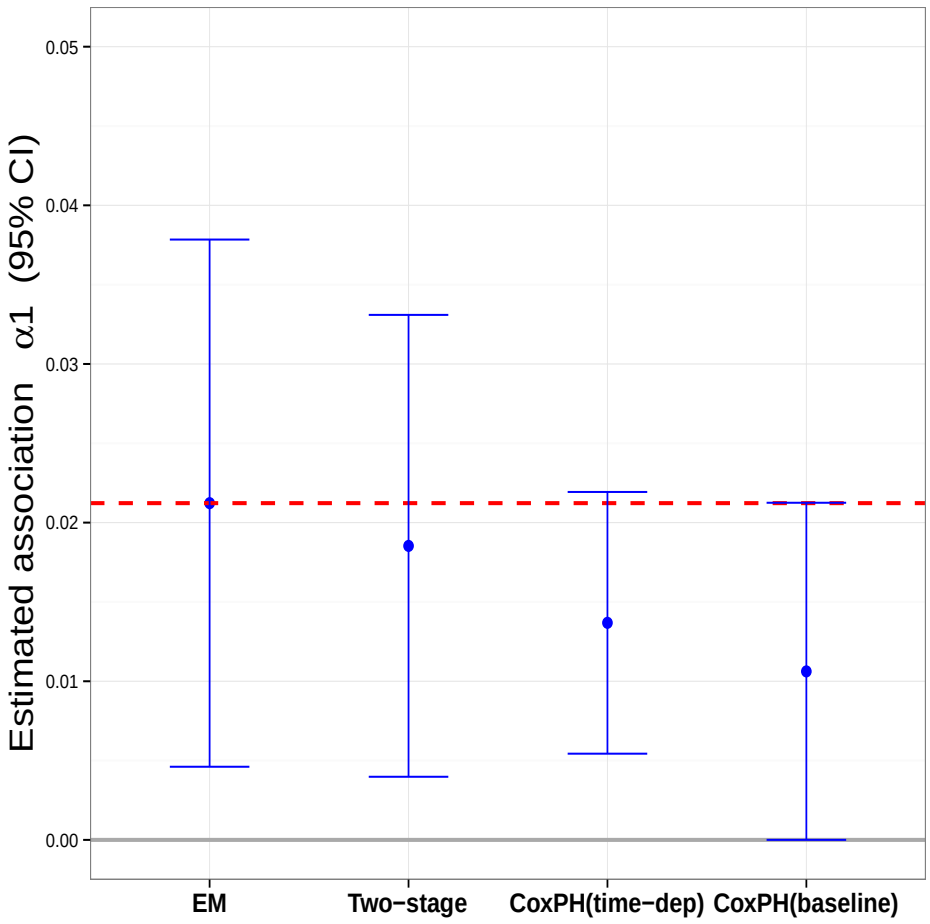

Figure 3.3: Comparison of estimated association ($\hat{\alpha_1}$) between the longitudinal systolic BP and risk of CAD from four methods. The blue solid dots are estimated $\hat{\alpha_1}$ from the four methods. The upper and lower bars are 95% CI of parameter estimates. The red dashed line denotes the estimate from the EM algorithm.

Table 3.8: Parameter estimates, standard errors and 95%CI for the joint Models 1. α_1 and α_2 are the association estimates between the risk of CAD and current value of systolic and diastolic BP at event time point, respectively. λ_i $i = 1, ..., 7$ denote the baseline hazards of the 7 piecewise constant intervals.

Parameter	Estimate	StdErr	lower 95%CI	upper 95%CI
Longitudinal Systolic BP				
Intercept	136.14	0.76	134.65	137.63
time	-0.25	0.08	-0.41	-0.10
Age	-0.00	0.05	-0.11	0.11
Race	4.54	0.80	2.97	6.12
$\log(\sigma_1)$	2.48	0.01	2.45	2.50
Longitudinal Diastolic BP				
Intercept	78.69	0.32	78.06	79.32
time	-1.07	0.04	-1.14	-0.99
Age	-0.12	0.02	-0.16	-0.08
Race	2.63	0.32	2.01	3.25
$\log(\sigma_1)$	1.95	0.01	1.92	1.97
Time-to-CAD				
Age	0.05	0.01	0.03	0.07
Smoking History	0.36	0.15	0.07	0.65
Race	-0.47	0.15	-0.77	-0.18
Diabetes	-0.03	0.14	-0.31	0.25
α_1	0.03	0.01	0.01	0.04
α_2	-0.004	0.01	-0.03	0.02
$\log(\lambda_1)$	-7.45	0.79	-9.01	-5.89
$\log(\lambda_2)$	-7.90	0.80	-9.46	-6.34
$\log(\lambda_3)$	-7.59	0.79	-9.14	-6.04
$\log(\lambda_4)$	-6.94	0.77	-8.46	-5.42
$\log(\lambda_5)$	-6.25	0.76	-7.74	-4.75
$\log(\lambda_6)$	-6.40	0.75	-7.88	-4.93
$\log(\lambda_7)$	-5.89	0.75	-7.36	-4.42

Table 3.9: Parameter estimates, standard errors and 95%CI for the joint Models 2. α_1 and α_2 are the association estimates between the risk of CAD and slope of systolic and diastolic BP at event time point, respectively. λ_i $i = 1, ..., 7$ denote the baseline hazards of the 7 piecewise constant intervals.

Parameter	Estimate	StdErr	lower 95%CI	upper 95%CI
Longitudinal Systolic BP				
Intercept	136.18	0.76	134.70	137.66
time	-0.26	0.08	-0.41	-0.10
Age	-0.00	0.05	-0.11	0.11
Race	4.49	0.80	2.92	6.06
$\log(\sigma_1)$	2.48	0.01	2.45	2.50
Longitudinal Diastolic BP				
Intercept	78.69	0.32	78.06	79.32
time	-1.06	0.04	-1.14	-0.99
Age	-0.12	0.02	-0.16	-0.07
Race	2.62	0.32	2.00	3.24
$\log(\sigma_1)$	1.95	0.01	1.92	1.97
Time-to-CAD				
Age	0.05	0.01	0.03	0.07
Smoking History	0.35	0.15	0.06	0.64
Race	-0.38	0.15	-0.67	-0.09
Diabetes	0.00	0.14	-0.27	0.28
α_1	0.41	0.19	0.04	0.78
α_2	-0.38	0.25	-0.87	0.11
$\log(\lambda_1)$	-4.26	0.35	-4.94	-3.58
$\log(\lambda_2)$	-4.70	0.36	-5.41	-4.00
$\log(\lambda_3)$	-4.38	0.35	-5.07	-3.69
$\log(\lambda_4)$	-3.74	0.32	-4.37	-3.10
$\log(\lambda_5)$	-3.06	0.31	-3.66	-2.46
$\log(\lambda_6)$	-3.22	0.30	-3.81	-2.64
$\log(\lambda_7)$	-2.71	0.30	-3.30	-2.12

Table 3.10: Parameter estimates, standard errors and 95%CI for the joint Models 3. α_1 and α_2 are the association estimates between the risk of CAD and current value of systolic and diastolic BP at event time point, respectively. λ_i $i = 1, ..., 7$ denote the baseline hazards of the 7 piecewise constant intervals.

Parameter	Estimate	StdErr	lower 95%CI	upper 95%CI
Longitudinal Systolic BP				
Intercept	135.53	0.80	133.95	137.10
time	0.26	0.16	-0.06	0.57
time^2	-0.03	0.01	-0.05	-0.01
Age	0.01	0.06	-0.10	0.11
Race	4.40	0.82	2.78	6.01
$\log(\sigma_1)$	2.47	0.01	2.45	2.50
Longitudinal Diastolic BP				
Intercept	79.42	0.34	78.75	80.09
time	-1.64	0.09	-1.82	-1.46
time^2	0.06	0.01	0.05	0.07
Age	-0.13	0.02	-0.18	-0.09
Race	2.74	0.32	2.11	3.36
$\log(\sigma_1)$	1.94	0.01	1.92	1.97
Time-to-CAD				
Age	0.06	0.01	0.04	0.08
Smoking History	0.35	0.15	0.06	0.65
Race	-0.49	0.15	-0.78	-0.19
Diabetes	-0.00	0.14	-0.28	0.28
α_1	0.021	0.008	0.005	0.038
α_2	0.011	0.014	-0.017	0.039
$\log(\lambda_1)$	-7.73	0.80	-9.30	-6.17
$\log(\lambda_2)$	-8.17	0.80	-9.74	-6.61
$\log(\lambda_3)$	-7.85	0.79	-9.41	-6.29
$\log(\lambda_4)$	-7.19	0.78	-8.72	-5.66
$\log(\lambda_5)$	-6.49	0.77	-8.00	-4.98
$\log(\lambda_6)$	-6.61	0.76	-8.10	-5.12
$\log(\lambda_7)$	-6.03	0.75	-7.50	-4.55

Table 3.11: Parameter estimates, standard errors and 95%CI for the joint Models 4. α_1 and α_2 are the association estimates between the risk of CAD and slope of systolic and diastolic BP at event time point, respectively. λ_i $i = 1, ..., 7$ denote the baseline hazards of the 7 piecewise constant intervals.

Parameter	Estimate	StdErr	lower 95%CI	upper 95%CI
Longitudinal Systolic BP				
Intercept	135.54	0.80	133.96	137.11
time	0.24	0.16	-0.07	0.55
time^2	-0.03	0.01	-0.05	-0.01
Age	0.00	0.06	-0.10	0.11
Race	4.40	0.82	2.79	6.02
$\log(\sigma_1)$	2.47	0.01	2.45	2.50
Longitudinal Diastolic BP				
Intercept	79.41	0.34	78.74	80.08
time	-1.64	0.09	-1.82	-1.46
time^2	0.06	0.01	0.05	0.07
Age	-0.13	0.02	-0.18	-0.09
Race	2.73	0.32	2.11	3.35
$\log(\sigma_1)$	1.94	0.01	1.92	1.97
Time-to-CAD				
Age	0.05	0.01	0.04	0.07
Smoking History	0.34	0.15	0.05	0.63
Race	-0.38	0.15	-0.67	-0.09
Diabetes	0.05	0.14	-0.23	0.32
α_1	0.11	0.17	-0.23	0.45
α_2	0.18	0.24	-0.30	0.65
$\log(\lambda_1)$	-3.73	0.48	-4.67	-2.79
$\log(\lambda_2)$	-4.19	0.44	-5.06	-3.32
$\log(\lambda_3)$	-3.89	0.39	-4.65	-3.13
$\log(\lambda_4)$	-3.28	0.31	-3.89	-2.67
$\log(\lambda_5)$	-2.63	0.25	-3.13	-2.14
$\log(\lambda_6)$	-2.83	0.21	-3.24	-2.42
$\log(\lambda_7)$	-2.38	0.26	-2.89	-1.86

3.8 Conclusion

We developed a maximum-likelihood method using the EM algorithm for parameter estimation of joint models for multiple longitudinal processes and a time-to-event outcome. Simulation studies indicated adequate performance of the EM based estimation approach which performed better than the two-stage estimation approach. We also applied the proposed method to data from a primary care patient cohort using EMR data for longitudinal systolic and diastolic BP and investigating their associations with the risk of CAD.

Our current work focused on joint models with normally distributed longitudinal outcomes. It is worth noting that the proposed methodology can be extended to joint models with other distributions for the longitudinal outcomes such as binary, Poisson and others. The proposed EM algorithm can be used for estimation from joint models with mixed types of longitudinal outcomes. Other potential extensions include compete-risk models or semi-compete-risk models to take informative censoring into consideration. Another area for further research is on predictive accuracy based on the proposed joint models.

The methodology for joint models of multiple longitudinal processes and time-to-event outcome is applicable to many clinical and epidemiologic studies where the association between longitudinal measures and time-to-event outcome is often of interest. The joint model framework provides a platform for exploring various features of the longitudinal measures related to disease risk, extending the traditional approach that relies on baseline measures only in cohort studies. With the increasing use of EMR in routine clinical practices, joint models can become a powerful tool for identifying longitudinal risk factors for disease risk and may offer insights for potential disease mechanisms that are otherwise not available using traditional approaches.

In clinical practice, for clinicians it may be also important to predict patients' survival probabilities based on their available multiple longitudinal biomarker measures. In addition,

the predictive ability of the longitudinal biomarkers in the joint modeling frameworks has received more and more attentions in the past few years. Rizopoulos (2011) and Njagi et al. (2013) have accessed the predictive ability of a single longitudinal biomarker in the joint modeling context. Adding more longitudinal biomarkers may improve the predictive ability of the risk model. Thus it is worthwhile to study the predictive ability of the multiple longitudinal biomarkers and to evaluate the improvement in the predictive performance by adding new longitudinal biomarkers in the joint modeling framework.

3.9 Acknowledgement

The research is supported by National Institutes of Health (NIH) Grants R01 AG019181, R24 MH080827, and P30 AG10133.

Chapter 4

Dynamic Predictions in Joint Models for Multiple Longitudinal Processes and Time-to-event Outcome

4.1 Abstract

In medical studies it is common to collect repeated biomarker measures over time along with the primary time-to-event outcome since these longitudinal biomarkers may be useful indicators and represent the disease progression. Joint models for longitudinal and survival data have been used to assess the association between the longitudinal outcomes and time-to-event outcome. Recently the predictive ability of the longitudinal outcome in joint models has also received a lot of attention. Recent literatures focus on the prediction in joint models with one single longitudinal outcome. However, the predictive ability of multiple longitudinal outcomes in joint models has not been studied even it is more common to collect multiple longitudinal biomarkers from participants. The question arises naturally - how much the prediction can be improved by adding new longitudinal biomarkers into the joint models? In this work, we extended existing approaches to predict conditional survival probabilities for joint models with multiple longitudinal biomarkers. We also applied novel prediction metrics to assess the improvement in prediction by adding new longitudinal outcome into the joint models. In addition, we compared the predictive performance of joint models to standard Cox models. Performance of proposed methods was assessed via simulations. The methodology was also applied to a real data set to predict the risk of coronary artery disease (CAD) using longitudinal systolic and diastolic blood pressures (BP).

4.2 Introduction

In Chapter 3, we developed a maximum-likelihood method using the EM algorithm for the parameter estimation of joint models for multiple longitudinal biomarkers and time-to-event outcome, where the main interest is to estimate the association between the multiple longitudinal variables and the risk of event. The repeated biomarker measures overtime are often useful indicators of the disease progression. The individualized prediction of joint models has also received an increasing attention in the past few years. In this chapter we focus on the following two aspects of predictions in joint models: to predict conditional survival probabilities in a clinical relevant time window, and to use the AUC (Hanley and McNeil, 1982) and other novel predictive accuracy criteria to evaluate the improved predictive ability by adding new longitudinal biomarkers into the joint models as well as the predictive performance comparison between joint models and standard Cox models.

There have been a few studies on prediction of survival probability in the joint modeling framework. Taylor et al. (2005) considered individualized predictions of disease progression using the joint models for a linear mixed-effect model and a logistic regression model. Yu et al. (2008) extended their previous work to the joint models for a linear mixed-effects model and a survival-cure model. Proust-Lima and Taylor (2009) focused on the disease recurrence prediction in the content of joint latent class model. Rizopoulos et al. (2013) developed a Bayesian model averaging approach for the prediction of the joint models. Taylor et al. (2013) predicted the probability of prostate cancer recurrence for a new patient using joint models and implemented on a web-based calculator. Rizopoulos (2011) studied an empirical Bayes approach and a MC simulation approach for the dynamic conditional survival probability predictions of the joint models for a single longitudinal outcome and a survival outcome. The empirical Bayes approach obtains the survival probability prediction by directly using the maximum-likelihood parameter estimates and individual prediction of

random effects, where the individual prediction of random effects is calculated by maximizing the posterior likelihood function of random effects. The empirical Bayes approach is computationally easy yet the derivation of the standard error for the survival probability is rather difficult. Therefore, a MC simulation approach was proposed by Rizopoulos et al. (2013). All these previous studies concentrated on predicting the survival probability of the joint models for a single longitudinal biomarker and a time-to-event outcome. To the best of our knowledge, we are not aware of studies which had looked into the survival probability predictions of the joint models for multiple longitudinal biomarkers and time-to-event outcome. We extended the empirical Bayes approach and the MC simulation approach to predict the conditional survival probability of joint models for multiple longitudinal biomarkers and a time-to-event outcome. The proposed methodology was applied to the longitudinal systolic and diastolic BP measures and time-to-CAD data from a primary care patient cohort.

Over the past decades, the assessment of predictive accuracy ability for survival analysis has received a lot of attention. The ROC (AUC) is the most popularly used discriminative criterion originally established for binary outcomes. Based on the concept of AUC for binary outcomes, several types of AUC were proposed for survival analysis. Pencina et al. (2012) provided a thorough review of existing AUCs for time-to-event outcomes. The most commonly used AUC for survival analysis was proposed by Harrell et al. (1982) and Harrell et al. (1996) and it has been further studied by Pencina and Agostino (2004). The AUC assesses the amount of concordance between predicted and observed outcomes comparing not only events and nonevents but also events that happened at different points in time. As an update, Uno et al. (2011) recently proposed a censoring-adjusted AUC based on the definition of Harrell et al. (1982). Chambless and Diao (2006) proposed a different time-dependent AUC focusing only on comparisons between event and nonevent. Gonen and

Heller (2005) considered a different way of assessing concordance applicable to proportional hazards models. Heagerty et al. (2000) proposed a dynamic ROC curve to summarize the discriminant of biomarker measured at baseline. Heagerty and Zheng (2005) further proposed ways to obtain estimate of time-dependent sensitivity, specificity and ROC curves based on the standard Cox regression model. Zheng and Heagerty (2007) and Antolini et al. (2005) extended to the survival analysis containing time-dependent covariates.

The AUC is a well-developed criterion for evaluating the discriminative ability for a single model. It has also been used for risk model comparison or evaluation of predictive ability by adding a new biomarker in medical studies. However, several studies showed that the AUC is insensitive in risk model comparison and the difference in AUC measures has no intuitive interpretation since it is only a function of rank but not predicted probabilities (Cook, 2007; Harrell, 2001; Janes et al., 2008; Moons and Harrell, 2003). Within the past few years, novel criteria have been proposed to quantify the improvement in model performance introduced by adding new biomarkers. Cook (2007) proposed a "reclassification table" to show how many subjects are reclassified if a new biomarker is added to the existing model. Pencina et al. (2008) extended the idea of reclassification table and proposed the Net Reclassification Improvement (NRI), and integrated discrimination improvement (IDI). IDI is a measure that integrates net reclassification over all possible cut-offs for the probability of the outcome. It is equivalent to the difference in discrimination slopes of two models (Yates, 1982), and to the difference in Pearson R^2 measures (Pepe et al., 2008), or the difference in scaled Brier scores (Gerds et al., 2008). Extensions have been made to account for time-to-event outcomes by Chambless et al. (2011) and Pencina et al. (2011). Uno et al. (2011) further updated to a more general type of NRI and IDI. Based on the definition of NRI and IDI, recently Pepe and Janes (2012) and Zheng et al. (2013) defined two new criteria, above average risk difference (AARD) and mean risk difference (MRD). They found

that the AARD is equivalent to the NRI for comparing a risk model to the model without any predictors. In addition, MRD between events and non-events is equivalent to the area between $\text{TPR}_t(p)$ and $\text{FPR}_t(p)$. Within the joint modelling framework, few studies have looked at the discriminative ability by using AUC (Njagi et al., 2013; Rizopoulos, 2011). Further more, no studies have investigated the predictive benefit by adding new longitudinal biomarkers in the joint modeling framework.

In this work, we explored the performance of AUC, AARD and MRD in evaluating the added predictive ability of a new longitudinal biomarker in the joint modeling framework via extensive simulations. In addition, we compared the predictive performance of joint models to standard Cox models. The comparison of the predictive performance of joint models to Cox time-dependent model is fair and straightforward since both models can incorporate the longitudinal data. However, when comparing the Cox baseline model to the joint models and Cox time-dependent model, one should notice that the Cox baseline model can only utilize the baseline measures of the longitudinal outcomes. Therefore, the joint models and Cox time-dependent model may be expected to show better prediction performance than the Cox baseline model. We also demonstrated the use of these criteria using data from a primary care patient cohort, as well as comparing the predictive performance of joint models to the aforementioned commonly used models.

The remainder of this chapter is organized as follows: Section 4.3 describes how the conditional survival probabilities can be estimated from the fitted joint model. Section 4.4 covers the definitions and estimators of the AUC, AARD and MRD in the joint modeling framework. Section 4.5 reports the results of simulation studies. In Section 4.6 we illustrated the proposed methodology to the data from a primary care patient cohort. Finally the chapter is concluded with a discussion in Section 4.7.

4.3 Predicting Conditional Survival Probabilities

In the last chapter, we focused on the statistical models and the parameter estimation method (using the EM algorithm) of the joint models for multiple longitudinal processes and time-to-event outcome. One important feature of this joint modeling framework is that the longitudinal biomarker trajectories are associated with the risk of event, implying that the longitudinal biomarker measures are directly related to the survival probabilities. Based on the maximum-likelihood estimates of the joint models and the multiple longitudinal measurements up to time t of a given new subject, one can predict the survival probability at any time point t. However, it may be more clinically relevant to predict the new subject's conditional survival probability at a future time point $t + \Delta t$ given the survival up to time t. Let $\mathcal{Y}_i(t) = \{y_{i1}(s), y_{i2}(s), ..., y_{Li}(s); 0 \leq s \leq t\}$ denote the ith subject's longitudinal biomarker measures up to time t for L different biomarkers and $\mathcal{D}_n = \{T_i, \delta_i, \mathbf{y}_{i1}, ..., \mathbf{y}_{iL}; i = 1, 2, ..., n\}$ represents the data set on which the joint models were fitted, the conditional survival probability at time $t + \Delta t$ given the survival up to time t can be written as

$$s_i(t + \Delta t|t) = P(T_i \geq t + \Delta t|T_i > t, \mathcal{Y}_i(t), \mathcal{D}_n; \boldsymbol{\theta}),$$

where T_i represents the new subject's observed event or censoring time. The conditional survival probability, $s_i(t + \Delta t|t)$, can be further decomposed as

$$\begin{aligned} s_i(t + \Delta t|t) &= P(T_i \geq t + \Delta t|T_i > t, \mathcal{Y}_i(t), \mathcal{D}_n; \boldsymbol{\theta}) \\ &= \int P(T_i \geq t + \Delta t|T_i > t, \mathcal{Y}_i(t), \mathbf{b}_i; \boldsymbol{\theta}) p(\mathbf{b}_i|T_i > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) d\mathbf{b}_i \\ &= \int \frac{S_i(t + \Delta t|\mathcal{Y}^*(t + \Delta t, \mathbf{b}_i, \boldsymbol{\theta}); \boldsymbol{\theta})}{S_i(t|\mathcal{Y}^*(t, \mathbf{b}_i, \boldsymbol{\theta}); \boldsymbol{\theta})} p(\mathbf{b}_i|T_i > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) d\mathbf{b}_i, \end{aligned}$$

where $\mathcal{Y}^*(t, \mathbf{b}_i, \boldsymbol{\theta})$ denotes the true multiple longitudinal biomarker measures. One straightforward approach to estimate $s_i(t + \Delta t|t)$ is to plug the empirical Bayes estimate of $\mathbf{b}_i$ ($\hat{\mathbf{b}}_i = \text{argmax}_{\mathbf{b}}\{\log p(T_i^* > t, \mathcal{Y}_i(t), \mathbf{b}; \hat{\boldsymbol{\theta}})\}$) and the maximum-likelihood parameter estimates ($\hat{\boldsymbol{\theta}}$) of the joint models into the above formula, leading to

$$\hat{s}_i(t + \Delta t|t) = \frac{S_i(t + \Delta t|\mathcal{Y}^*(t + \Delta t, \hat{\mathbf{b}_\mathbf{i}}, \hat{\boldsymbol{\theta}}); \hat{\theta})}{S_i(t|\mathcal{Y}^*(t, \hat{\mathbf{b}_\mathbf{i}}, \hat{\boldsymbol{\theta}}); \hat{\theta})} + O(n_i^{-1}).$$

This empirical Bayes approach has the advantage of easy computation, but the derivation of its standard error is complicate and not straightforward. Rizopoulos (2011) proposed the MC simulation approach to predict conditional survival probabilities. Basically, instead of directly calculating the conditional survival probability $s_i(t+\Delta t|t)$, the predicted conditional survival probability is summarized from a series of posterior expectation of $s_i(t+\Delta t|t)$ which is formulated as follows:

$$\begin{aligned} P(T_i \geq t + \Delta t|T_i > t, \mathcal{Y}_i(t), \mathcal{D}_n) &= \int P(T_i \geq t + \Delta t|T_i > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}_n) d\boldsymbol{\theta} \\ &= \int \int \frac{S_i(t + \Delta t|\mathcal{Y}^*(t + \Delta t, \mathbf{b}_i, \boldsymbol{\theta}); \boldsymbol{\theta})}{S_i(t|\mathcal{Y}^*(t, \mathbf{b}_i, \boldsymbol{\theta}); \boldsymbol{\theta})} p(\boldsymbol{\theta}|\mathcal{D}_n) \\ &\quad p(\mathbf{b}_i|T_i > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{b}_i. \end{aligned}$$

The prediction of $s_i(t + \Delta t|t)$ using the MC simulation approach can be achieved from the following steps:

Step 1: Randomly simulate $\boldsymbol{\theta}^{(m)}$ from a normal distribution $\mathcal{N}(\hat{\boldsymbol{\theta}}, \hat{\text{var}}(\hat{\boldsymbol{\theta}}))$. Here we use the asymptotic Bayesian theory (Cox, 1972) and assume that the sample size n is sufficiently large such that $\{\theta|\mathcal{D}_n\}$ can be well approximated by $\mathcal{N}(\hat{\boldsymbol{\theta}}, \hat{\text{var}}(\hat{\boldsymbol{\theta}}))$.

Step 2: Randomly simulate $\mathbf{b}_i^{(m)}$ from a multivariate t distribution centered at the empirical

Bayes estimate $\hat{\mathbf{b}_i}$, with scale matrix

$$\hat{\text{var}}(\hat{\mathbf{b}_i}) = \{\frac{-\partial^2 \log p(T_i > t, \mathcal{Y}_i(t), \mathbf{b}; \hat{\boldsymbol{\theta}})}{\partial \mathbf{b}^T \partial \mathbf{b}}|_{\mathbf{b}=\hat{\mathbf{b}_i}}\}^{-1}$$

and four degrees of freedom.

Step 3: Calculate

$$s_i^{(m)}(t + \Delta t|t) = \frac{S_i(t + \Delta t|\mathcal{Y}^*(t + \Delta t, \mathbf{b}_i^{(m)}, \boldsymbol{\theta}^{(m)}); \boldsymbol{\theta}^{(m)})}{S_i(t|\mathcal{Y}^*(t, \mathbf{b}_i^{(m)}, \boldsymbol{\theta}^{(m)}); \boldsymbol{\theta}^{(m)})}$$

Step 4: Repeat Step 1 to Step 3 for all the subjects, where $m = 1, ..., M$ denotes the number of samples.

Step 5: The mean or median of $\{s_i^{(m)}(t + \Delta t|t), m = 1, ..., M\}$ is used as the estimate of $s_i(t + \Delta t|t)$. The square root of the sample variance over $\{s_i^{(m)}(t + \Delta t|t), m = 1, ..., M\}$ is used as the standard error.

It is noted that the conditional survival probability prediction can be progressively updated as more longitudinal biomarker measurements become available. Such dynamic predictions can be useful in clinical practice.

4.4 Predictive Accuracy

The predictive accuracy of a survival model includes calibration, discrimination and reclassification. Calibration focuses on quantifying how close the predicted outcomes are to the observed outcomes and the discrimination is to quantify how well a model can distinguish subjects with event from those who won't experience event, while reclassification is to evaluate the incremental values from using new predictors in the model. In this work we concentrate on the ability of reclassification of new longitudinal biomarker in the joint modeling framework. Here we propose to define the TPR and FPR using a subject's risk

by a future time t, i.e.,

$$r_i(t) = P(T_i \leq t | \mathcal{Y}_i^*(t), \mathcal{D}_n; \boldsymbol{\theta}) = 1 - S_i(t | \mathcal{Y}^*(t, \mathbf{b}_i, \boldsymbol{\theta}); \boldsymbol{\theta}),$$

where the risk at time t, $r_i(t)$, depends on the predicted longitudinal biomarker measures up to time t and the fitted joint models. TPR and FPR for a given threshold p are defined as follows:

$$\begin{aligned} \text{TPR}_t(p) &= P(\mathbf{r}(t) \geq p | \mathbf{T} \leq t), \\ \\ \text{FPR}_t(p) &= P(\mathbf{r}(t) \geq p | \mathbf{T} > t). \end{aligned}$$

Commonly used estimators of TPR and FPR are defined as in the following,

$$\widehat{\text{TPR}}_t(p) = \frac{\sum_{i=1}^{n^*} I(\hat{r}_i(t) \geq p) I(T_i \leq t)}{\sum_{i=1}^{n^*} I(T_i \leq t)}, \tag{4.1}$$

$$\widehat{\text{FPR}}_t(p) = \frac{\sum_{i=1}^{n^*} I(\hat{r}_i(t) \geq p) I(T_i > t)}{\sum_{i=1}^{n^*} I(T_i > t)}, \tag{4.2}$$

where n^* denotes the number of subjects experiencing events. Zheng et al. (2013) proposed another type of estimator by incorporating both censors and events,

$$\widehat{\text{TPR}}_t(p) = \frac{\sum_{i=1}^{n} \hat{r}_i(t) I(\hat{r}_i(t) \geq p)}{\sum_{i=1}^{n} \hat{r}_i(t)}, \tag{4.3}$$

$$\widehat{\text{FPR}}_t(p) = \frac{\sum_{i=1}^{n} (1 - \hat{r}_i(t)) I(\hat{r}_i(t) \geq p)}{\sum_{i=1}^{n} (1 - \hat{r}_i(t))}. \tag{4.4}$$

Based on the definition of the AUC, the AUC at time t can be written as

$$\text{AUC}_t = \int \text{TPR}_t \text{FPR}_t^{-1}(u)du.$$

Using the estimates of FPR and TPR, the estimation of AUC can be readily derived,

$$\widehat{\text{AUC}}_t = \int \widehat{\text{TPR}}_t \widehat{\text{FPR}}_t^{-1}(u)du.$$

The newly proposed AARD and MRD (Pepe and Janes, 2012; Zheng et al., 2013) can also be defined by the risk function, $r(t)$:

$$\text{AARD} = P(\mathbf{r}(t) > \rho_t | \mathbf{T} \leq t) - P(\mathbf{r}(t) > \rho_t | \mathbf{T} > t),$$

$$\text{MRD} = E(\mathbf{r}(t) | \mathbf{T} \leq t) - E(\mathbf{r}(t) | \mathbf{T} > t),$$

where $\rho_t = P(\mathbf{T} \leq t)$. Similarly, the estimation of AARD and MRD can be written as follows:

$$\widehat{\text{AARD}}_t = \widehat{\text{TPR}}_t(\hat{\rho}) - \widehat{\text{FPR}}_t(\hat{\rho}),$$

$$\widehat{\text{MRD}}_t = \int_p \widehat{\text{TPR}}_t(p)dp - \int_p \widehat{\text{FPR}}_t(p)dp,$$

where

$$\hat{\rho} = \frac{1}{\sum_{i=1}^n \hat{r}_i(t)}.$$

4.5 Simulation Study

We used simulations to assess performance of the empirical Bayes approach and the MC simulation approach in predicting conditional survival probabilities of joint models for multiple longitudinal biomarkers and a time-to-event outcome. Performance of aforementioned

three criteria for evaluating the improvement in predictive ability by adding new biomarkers was also investigated via extensive simulations.

The joint models described in the simulation section of the previous chapter were used to generate data. Three simulation scenarios were considered by varying variances of residual errors and variance-covariance matrix of random effects. The three scenarios are presented in Table 4.1. All other true parameter values were summarized in Table 4.2. Two hundred training data sets were generated for each scenario. Each training data set consists of 500 subjects with each subject having up to 11 equally spaced bivariate longitudinal evaluations over a 5 year period. Additional 200 testing data sets were simulated using the same covariate values and random effects as in training data sets for each scenario .

Table 4.1: Three scenarios differing in variances of residual errors and variances of random effects used in simulations.

Parameter	Scenario 1	Scenario 2	Scenario 3
σ_1	0.2	0.2	0.5
σ_2	0.2	0.2	0.5
σ_{01}	0.2	0.5	0.5
σ_{11}	0.05	0.2	0.2
σ_{02}	0.2	0.5	0.5
σ_{12}	0.05	0.2	0.2

Table 4.2: Other true parameter values for the two longitudinal models and the Cox PH model used in simulations.

Outcome 1		Outcome 2		Time-to-event outcome	
Parameter	Value	Parameter	Value	Parameter	Value
β_{01}	0.2	β_{02}	1.0	a	0.005
β_{11}	0.5	β_{12}	0.2	b	1.1
β_{21}	0.2	β_{22}	0.5	α_1	1.0
ρ	0.5			α_2	1.0

4.5.1 Predicting Conditional Survival Probabilities

We fitted joint models to each training data set and estimated parameters using the EM algorithm. For each testing data set, data from the first 2 years was used to predict individual random effects. Twenty subjects were randomly selected from each testing data set. Selected subjects' conditional survival probabilities at different times t ($t > 2$) and Δt were predicted via the empirical Bayes approach and the MC simulation approach, using the estimated parameters from training data set and predicted individual random effects from the first 2 year testing data. In the MC simulation approach, the median based on 200 MC replicates was used as the predicted conditional survival probability. Predicted conditional survival probabilities obtained from the two approaches were compared to the true predictions based on the true random effects and true parameter values. We reported the average biases between the true and predicted conditional survival probabilities for the selected subjects over the 200 testing data sets.

Simulation results for the three scenarios are presented in Tables 4.3, 4.4 and 4.5. From simulation results, it is evident that for scenario 1 (small variances of residual errors and small variances of random effects), the performance of the empirical Bayes approach is fairly comparable to that of the MC simulation approach: the biases in conditional survival probability predictions are ignorable. However, under the large variance scenarios 2 and 3, the MC simulation approach generally performs better in predicting conditional survival probabilities than the empirical Bayes approach - smaller biases are observed for the MC simulation approach. In addition, the results clearly show that for fixed Δt, biases increase as time t increases. This may be explained by the fact that more subjects drop out as time t increases leading to less accurate random effect predictions. On the other hand, the results also indicate that for fixed time t, biases generally increase as the prediction window

Δt increases. In other words, conditioning on the same number of longitudinal biomarker measures, the performance of prediction gets worse when window for prediction is longer.

4.5.2 Predictive Accuracy

Additional simulations were conducted to examine the performance of AUC, AARD and MRD in evaluating the predictive ability of joint models. The predictive performance of joint models were also compared to other commonly used models. In each training data set we fitted the following four models: the Cox PH model with longitudinal biomarker measures at baseline as time-independent covariates (Cox baseline), Cox PH model with longitudinal biomarker measures as time-dependent (Cox time-dependent), joint models of the first longitudinal biomarker and time-to-event outcome (JM1), and joint models of bivariate longitudinal biomarkers and time-to-event outcome (JM2). For JM1 and JM2 models, individual survival probabilities at time t ($t > 2$) for subjects from testing data set were obtained using the similar procedures as described for predicting conditional survival probabilities in the previous section. For the Cox baseline model, individual survival probabilities were calculated using the estimated parameters from the training data set and longitudinal measures at baseline from the testing data. The same technique was applied for the Cox time-dependent model, except for using the longitudinal measures up to time t from the testing data. Based on predicted survival probabilities and estimators for TPR and FPR (formulas (4.1) and (4.2)), TPR and FPR can be readily calculated for different models. Accordingly, AUC, AARD and MRD were derived from these TPR and FPR values.

The sample means and empirical standard errors of AUC, AARD, and MRD for the four models were summarized over the 200 testing data sets for each scenario. Simulation results with respect to different scenarios are presented in Tables 4.6, 4.7 and 4.8. It is shown that

among the four models, JM2 has the best predictive performance - little differences were observed between the true predictive accuracy criteria (calculated based on true random effects and true parameters) and the calculated predictive accuracy criteria based on JM2. From results of JM2 we observe that the predictive performance improves as the time t increases. It is also evident that the variances of residual of errors and random effects have some impact on the predictive performance: larger variances of residual of errors lead to larger biases (Tables 4.7 and 4.8); larger variances of random effects also result in larger biases (Tables 4.6 and 4.7).

Simulation results of JM1 and JM2 show that for large prediction window t, AUC seems be able to reflect the improvement of predictive ability by adding the second longitudinal biomarker. However, when the prediction window t is small, little improvement was observed in AUC. The cause of such phenomenon may be that for large prediction window t the effect on the survival from the second longitudinal outcome may be large enough for AUC to capture the predictive improvement; whereas for small prediction window t, the effect from the second longitudinal outcome is not large enough for AUC to reflect the predictive improvement. This observed phenomenon of insensitivity of AUC is similar to discussions in Janes et al. (2008). Compared to AUC, AARD and MRD seem to be more sensitive in measuring the improvement of predictive ability from the second longitudinal biomarker.

Simulation results also reveal persistent biases between true predictive criteria (calculated based on true random effects and true parameter values) and calculated criteria from JM2. We notice that AUC, AARD, and MRD rely on predicted survival probabilities. Thus the precision of parameter estimates and the random effect predictions are directly related to precision of predictive accuracy criteria. Extra simulations were conducted to determine how the accuracy of random effect predictions and parameter estimates influence the per-

formance of the predictive criteria. We defined two types of pseudo survival probability estimators: using true random effects and estimated parameter values (Pseudo 1) and using predicted random effects and true parameter values (Pseudo 2). Simulation results are presented in Tables 4.9, 4.10 and 4.11. It is observed that both the accuracy of parameter estimates and random effect predictions have impact on the precision of predictive criteria.

Table 4.3: Biases for comparing predicted conditional survival probabilities of the empirical Bayes approach to the MC simulation approach under Scenario 1. For the MC simulation approach, the median of predicted conditional survival probabilities over the 200 MC draws was used. The values in the bracket are the lower 2.5% and upper 97.5% percentile of the predictions from all testing data sets.

	$\Delta t = 2.5$	$\Delta t = 3$	$\Delta t = 3.5$	$\Delta t = 4$
			MC simulation approach	
$t = 2.5$	-0.002 (-0.16, 0.145)	-0.005 (-0.161, 0.133)	-0.007 (-0.149, 0.101)	-0.007 (-0.12, 0.069)
$t = 3$	-0.005 (-0.164, 0.137)	-0.007 (-0.149, 0.106)	-0.007 (-0.126, 0.071)	-0.006 (-0.095, 0.041)
$t = 3.5$	-0.006 (-0.156, 0.116)	-0.007 (-0.129, 0.082)	-0.006 (-0.098, 0.045)	-0.003 (-0.06, 0.022)
$t = 4$	-0.006 (-0.137, 0.089)	-0.006 (-0.104, 0.05)	-0.004 (-0.062, 0.023)	-0.002 (-0.029, 0.008)
			empirical Bayes approach	
$t = 2.5$	-0.002 (-0.169, 0.152)	-0.005 (-0.168, 0.144)	-0.007 (-0.156, 0.111)	-0.007 (-0.131, 0.078)
$t = 3$	-0.005 (-0.174, 0.15)	-0.007 (-0.16, 0.116)	-0.007 (-0.137, 0.082)	-0.006 (-0.097, 0.046)
$t = 3.5$	-0.008 (-0.168, 0.122)	-0.008 (-0.143, 0.087)	-0.006 (-0.102, 0.049)	-0.004 (-0.063, 0.021)
$t = 4$	-0.009 (-0.152, 0.094)	-0.007 (-0.11, 0.054)	-0.004 (-0.069, 0.023)	-0.002 (-0.033, 0.007)

Table 4.4: Biases for comparing predicted conditional survival probability of the empirical Bayes approach to the simulation approach under Scenario 2. For the MC simulation approach, the median of predicted conditional survival probabilities over the 200 MC draws was used. The values in the bracket are the lower 2.5% and upper 97.5% percentile of the predictions from all testing data sets.

	$\Delta t = 2.5$	$\Delta t = 3$	$\Delta t = 3.5$	$\Delta t = 4$
			MC simulation approach	
$t = 2.5$	-0.003 (-0.287, 0.261)	-0.011 (-0.319, 0.221)	-0.016 (-0.331, 0.181)	-0.019 (-0.313, 0.149)
$t = 3$	-0.007 (-0.325, 0.248)	-0.014 (-0.338, 0.198)	-0.017 (-0.321, 0.161)	-0.017 (-0.314, 0.138)
$t = 3.5$	-0.009 (-0.333, 0.228)	-0.015 (-0.317, 0.174)	-0.017 (-0.31, 0.149)	-0.017 (-0.293, 0.125)
$t = 4$	-0.008 (-0.319, 0.199)	-0.014 (-0.314, 0.159)	-0.015 (-0.292, 0.133)	-0.015 (-0.279, 0.114)
			empirical Bayes approach	
$t = 2.5$	-0.006 (-0.336, 0.294)	-0.014 (-0.373, 0.277)	-0.019 (-0.383, 0.259)	-0.022 (-0.372, 0.238)
$t = 3$	-0.014 (-0.386, 0.287)	-0.02 (-0.39, 0.27)	-0.023 (-0.384, 0.244)	-0.023 (-0.37, 0.217)
$t = 3.5$	-0.021 (-0.405, 0.281)	-0.024 (-0.396, 0.253)	-0.024 (-0.386, 0.225)	-0.023 (-0.377, 0.2)
$t = 4$	-0.025 (-0.42, 0.265)	-0.025 (-0.401, 0.236)	-0.024 (-0.383, 0.205)	-0.022 (-0.373, 0.175)

Table 4.5: Biases for comparing predicted conditional survival probability of the empirical Bayes approach to the simulation approach under Scenario 3. For the MC simulation approach, the median of predicted conditional survival probabilities over the 200 MC draws was used. The values in the bracket are the lower 2.5% and upper 97.5% percentile of the predictions from all testing data sets.

	$\Delta t = 2.5$	$\Delta t = 3$	$\Delta t = 3.5$	$\Delta t = 4$
		MC simulation approach		
$t = 2.5$	0.002 (-0.362, 0.361)	-0.014 (-0.403, 0.296)	-0.024 (-0.423, 0.254)	-0.03 (-0.424, 0.214)
$t = 3$	-0.006 (-0.398, 0.321)	-0.02 (-0.427, 0.267)	-0.027 (-0.427, 0.23)	-0.03 (-0.414, 0.191)
$t = 3.5$	-0.01 (-0.425, 0.294)	-0.023 (-0.444, 0.249)	-0.028 (-0.434, 0.202)	-0.029 (-0.418, 0.164)
$t = 4$	-0.011 (-0.419, 0.287)	-0.022 (-0.415, 0.228)	-0.026 (-0.416, 0.174)	-0.026 (-0.412, 0.139)
		empirical Bayes approach		
$t = 2.5$	-0.004 (-0.405, 0.384)	-0.019 (-0.447, 0.344)	-0.029 (-0.465, 0.316)	-0.035 (-0.456, 0.278)
$t = 3$	-0.019 (-0.461, 0.357)	-0.03 (-0.476, 0.318)	-0.036 (-0.473, 0.289)	-0.037 (-0.478, 0.243)
$t = 3.5$	-0.031 (-0.499, 0.332)	-0.037 (-0.492, 0.298)	-0.039 (-0.488, 0.252)	-0.037 (-0.475, 0.19)
$t = 4$	-0.039 (-0.513, 0.315)	-0.041 (-0.497, 0.264)	-0.039 (-0.489, 0.197)	-0.035 (-0.453, 0.148)

Table 4.6: Comparison of AUC, AARD, and MRD from JM2 to the other 3 models under simulation scenario 1.

		Cox(Baseline)		Cox(time-dependent)		JM1		JM2	
t	True	Mean	ESD	Mean	ESD	Mean	ESD	Mean	ESD
					AUC				
3	0.659	0.634	0.028	0.501	0.031	0.624	0.029	0.649	0.028
3.5	0.673	0.643	0.026	0.497	0.029	0.634	0.028	0.662	0.027
4	0.692	0.657	0.026	0.497	0.028	0.644	0.027	0.679	0.027
4.5	0.715	0.672	0.026	0.496	0.029	0.657	0.025	0.699	0.025
5	0.745	0.691	0.027	0.499	0.032	0.676	0.026	0.725	0.025
					AARD				
3	0.230	0.198	0.048	0.013	0.054	0.180	0.051	0.219	0.048
3.5	0.252	0.210	0.045	-0.002	0.050	0.194	0.051	0.237	0.048
4	0.281	0.232	0.045	-0.018	0.049	0.211	0.047	0.262	0.048
4.5	0.316	0.255	0.043	-0.036	0.052	0.226	0.047	0.293	0.048
5	0.363	0.283	0.048	-0.043	0.055	0.254	0.047	0.331	0.044
					MRD				
3	0.065	0.044	0.015	0.003	0.013	0.028	0.009	0.055	0.014
3.5	0.088	0.058	0.018	0.000	0.016	0.040	0.011	0.075	0.017
4	0.113	0.073	0.020	-0.002	0.018	0.052	0.013	0.096	0.020
4.5	0.134	0.084	0.022	-0.004	0.019	0.062	0.014	0.114	0.021
5	0.147	0.088	0.022	-0.002	0.020	0.069	0.014	0.125	0.021

Table 4.7: Comparison of AUC, AARD, and MRD from JM2 to the other 3 models under simulation scenario 2.

		Cox(Baseline)		Cox(time-dependent)		JM1		JM2	
t	True	Mean	ESD	Mean	ESD	Mean	ESD	Mean	ESD
					AUC				
3	0.791	0.738	0.026	0.683	0.027	0.720	0.025	0.774	0.023
3.5	0.811	0.737	0.023	0.686	0.024	0.728	0.023	0.790	0.021
4	0.830	0.736	0.022	0.688	0.023	0.733	0.022	0.805	0.019
4.5	0.851	0.734	0.022	0.692	0.022	0.738	0.023	0.819	0.018
5	0.871	0.735	0.022	0.697	0.023	0.745	0.023	0.834	0.017
					AARD				
3	0.434	0.348	0.048	0.263	0.048	0.323	0.047	0.407	0.046
3.5	0.465	0.345	0.043	0.267	0.042	0.335	0.043	0.433	0.041
4	0.502	0.343	0.042	0.264	0.042	0.342	0.043	0.454	0.039
4.5	0.540	0.341	0.042	0.267	0.042	0.347	0.042	0.478	0.039
5	0.582	0.343	0.044	0.270	0.044	0.355	0.042	0.503	0.038
					MRD				
3	0.258	0.130	0.026	0.161	0.028	0.088	0.018	0.203	0.025
3.5	0.303	0.146	0.027	0.175	0.027	0.108	0.021	0.241	0.025
4	0.337	0.153	0.026	0.173	0.026	0.125	0.023	0.270	0.025
4.5	0.362	0.154	0.026	0.164	0.026	0.136	0.023	0.292	0.024
5	0.379	0.149	0.025	0.155	0.026	0.145	0.024	0.306	0.024

Table 4.8: Comparison of AUC, AARD, and MRD from JM2 to the other 3 models under simulation scenario 3.

		Cox(Baseline)		Cox(time-dependent)		JM1		JM2	
t	True	Mean	ESD	Mean	ESD	Mean	ESD	Mean	ESD
					AUC				
3	0.791	0.696	0.027	0.659	0.028	0.696	0.025	0.749	0.024
3.5	0.811	0.696	0.025	0.663	0.025	0.704	0.023	0.763	0.021
4	0.830	0.694	0.022	0.665	0.024	0.709	0.022	0.775	0.019
4.5	0.851	0.693	0.023	0.669	0.023	0.713	0.023	0.786	0.019
5	0.871	0.694	0.024	0.675	0.024	0.717	0.022	0.798	0.019
					AARD				
3	0.434	0.288	0.048	0.231	0.048	0.288	0.045	0.368	0.050
3.5	0.465	0.285	0.044	0.233	0.044	0.299	0.043	0.388	0.041
4	0.502	0.281	0.042	0.229	0.043	0.303	0.045	0.405	0.038
4.5	0.540	0.278	0.042	0.231	0.044	0.308	0.046	0.420	0.039
5	0.582	0.279	0.043	0.233	0.045	0.311	0.042	0.436	0.038
					MRD				
3	0.258	0.084	0.022	0.112	0.024	0.071	0.016	0.154	0.021
3.5	0.303	0.097	0.023	0.127	0.024	0.089	0.019	0.186	0.022
4	0.337	0.103	0.023	0.127	0.023	0.103	0.020	0.211	0.022
4.5	0.362	0.104	0.023	0.119	0.022	0.114	0.021	0.230	0.022
5	0.379	0.102	0.022	0.108	0.022	0.121	0.021	0.244	0.022

Table 4.9: Simulation results for comparing AUC, AARD, and MRD for the three different survival probability estimators under scenario 1. Pseudo 1 denotes the estimator using true random effects and estimated parameter values; Pseudo 2 denotes the estimator using estimated random effects and true parameter values; JM2 denotes the estimator using estimated random effects and estimated parameters.

		Pseudo 1		Pseudo 2		JM2	
t	True	Mean	ESD	Mean	ESD	Mean	ESD
				AUC			
3	0.659	0.659	0.027	0.650	0.028	0.649	0.028
3.5	0.673	0.673	0.026	0.664	0.027	0.662	0.027
4	0.692	0.692	0.026	0.682	0.027	0.679	0.027
4.5	0.715	0.715	0.025	0.701	0.025	0.699	0.025
5	0.745	0.745	0.025	0.728	0.025	0.725	0.025
				AARD			
3	0.230	0.231	0.046	0.218	0.048	0.219	0.048
3.5	0.252	0.252	0.045	0.238	0.049	0.237	0.048
4	0.281	0.282	0.045	0.265	0.048	0.262	0.048
4.5	0.316	0.315	0.046	0.297	0.045	0.293	0.048
5	0.363	0.364	0.043	0.335	0.045	0.331	0.044
				MRD			
3	0.065	0.067	0.016	0.052	0.012	0.055	0.014
3.5	0.088	0.092	0.020	0.072	0.013	0.075	0.017
4	0.113	0.117	0.024	0.093	0.013	0.096	0.020
4.5	0.134	0.137	0.025	0.110	0.015	0.114	0.021
5	0.147	0.151	0.026	0.123	0.022	0.125	0.021

Table 4.10: Simulation results for comparing AUC, AARD, and MRD for the three different survival probability estimators under scenario 2. Pseudo 1 denotes the estimator using true random effects and estimated parameter values; Pseudo 2 denotes the estimator using estimated random effects and true parameter values; JM2 denotes the estimator using estimated random effects and estimated parameter values.

		Pseudo 1		Pseudo 2		JM2	
t	True	Mean	ESD	Mean	ESD	Mean	ESD
				AUC			
3	0.791	0.791	0.023	0.780	0.023	0.774	0.023
3.5	0.811	0.811	0.021	0.800	0.021	0.790	0.021
4	0.830	0.830	0.018	0.817	0.019	0.805	0.019
4.5	0.851	0.851	0.016	0.835	0.017	0.819	0.018
5	0.871	0.871	0.016	0.853	0.017	0.834	0.017
				AARD			
3	0.434	0.433	0.048	0.416	0.048	0.407	0.046
3.5	0.465	0.466	0.045	0.445	0.042	0.433	0.041
4	0.502	0.501	0.039	0.474	0.040	0.454	0.039
4.5	0.540	0.540	0.037	0.505	0.039	0.478	0.039
5	0.582	0.583	0.039	0.538	0.038	0.503	0.038
				MRD			
3	0.258	0.259	0.031	0.222	0.024	0.203	0.025
3.5	0.303	0.303	0.031	0.267	0.022	0.241	0.025
4	0.337	0.335	0.029	0.302	0.022	0.270	0.025
4.5	0.362	0.356	0.028	0.328	0.025	0.292	0.024
5	0.379	0.368	0.028	0.344	0.033	0.306	0.024

Table 4.11: Simulation results for comparing AUC, AARD, and MRD for the three different survival probability estimators under scenario 3. Pseudo 1 denotes the estimator using true random effects and estimated parameter values; Pseudo 2 denotes the estimator using estimated random effects and true parameter values; JM2 denotes the estimator using estimated random effects and estimated parameter values.

		Pseudo 1		Pseudo 2		JM2	
t	True	Mean	ESD	Mean	ESD	Mean	ESD
				AUC			
3	0.791	0.791	0.023	0.757	0.024	0.749	0.024
3.5	0.811	0.810	0.021	0.777	0.022	0.763	0.021
4	0.830	0.830	0.018	0.793	0.020	0.775	0.019
4.5	0.851	0.851	0.016	0.809	0.018	0.786	0.019
5	0.871	0.871	0.016	0.824	0.018	0.798	0.019
				AARD			
3	0.434	0.432	0.047	0.380	0.049	0.368	0.050
3.5	0.465	0.465	0.044	0.408	0.044	0.388	0.041
4	0.502	0.501	0.038	0.434	0.039	0.405	0.038
4.5	0.540	0.539	0.037	0.461	0.038	0.420	0.039
5	0.582	0.582	0.039	0.486	0.040	0.436	0.038
				MRD			
3	0.258	0.260	0.032	0.180	0.025	0.154	0.021
3.5	0.303	0.304	0.032	0.225	0.024	0.186	0.022
4	0.337	0.335	0.030	0.259	0.022	0.211	0.022
4.5	0.362	0.357	0.029	0.285	0.024	0.230	0.022
5	0.379	0.369	0.029	0.299	0.032	0.244	0.022

4.6 Data Application to A Primary Care Patient Cohort

We first applied joint models to the primary care patient cohort data discussed in Chapter 3. Out of the 797 subjects, we randomly selected 597 subjects to create the testing data set and the remaining 200 subjects comprised the testing data set. For convenience we centered patients' baseline age at 60 years. We fitted four different sets of joint models introduced in Chapter 3 Section 3.7 to the training data set using the proposed EM algorithm. In the implementation of the EM algorithm, we used 3 pseudo-adaptive Gaussian-Hermite quadrature points for numerical integration over the random effects and 7 Gaussian-Kronrod quadrature points for the integration in the survival function. The best set of models were determined using the AIC: smaller AIC indicates better model fit. Among the 4 joint models considered, Joint models 3 was the best fitting (AIC=49464) followed by Joint models 4 (AIC=49475), Joint models 1 (AIC=49537) and Joint models 2 (AIC=49548). Here we presented parameter estimates from joint models 3 in Table 4.12. The results imply a similar conclusion as in the data application section of Chapter 3: a high systolic BP measure is significantly associated with a high risk of CAD, while a non-significant association with risk of CAD is observed for the diastolic BP measure.

We then used data from the first 5 years in the testing data set to estimate patients' subject-specific random effects. Based on parameter estimates from the training data set and the random effect estimates from the first 5 year testing data, predicted conditional survival probabilities given t ($t > 5$) were further computed using the MC simulation approach. In the following we assessed the performance of predictive ability of the joint models using the testing data set, as well as compared the prediction performance of joint models with systolic and diastolic BP measures to joint models with one single type of BP measure (systolic or diastolic) and other standard Cox models.

4.6.1 Predicting Conditional Survival Probabilities

In this section, we focused on predicting conditional survival probabilities in joint models. As an example, we chose two subjects from the testing data set to illustrate how the longitudinal BP measures over time influence the conditional survival probability predictions. We selected subject 143 and 318 with the same baseline risk covariates. Subject 143 was a 66 years old black male with a history of smoking and diabetes, and was lost to follow up 6.97 years after baseline. Subject 318 had the same demographics as subject 143, except that a CAD event was observed at year 7.5. The two selected subjects with the same characteristics allowed us to study the pure affect of longitudinal BP measures over time on the risk of developing CAD. The longitudinal BP measures over time for the two subjects were plotted in Figure 4.1. It is observed that, in general, longitudinal BP measures of subject 143 increased and then decreased over time, while subject 318 had an increasing and then decreasing trend in BP measures over time. The two subjects' predicted conditional survival probabilities were summarized in Table 4.13 and Figures 4.2 and 4.3. From the two plots, we clearly observed how the longitudinal BP measures impacted on the risk of developing CAD. Overall, subject 318 had larger risk in developing CAD than subject 143.

4.6.2 Predictive Accuracy

We assessed the performance of predictive ability of the joint models with longitudinal systolic and diastolic BP measures (JM2) using the testing data set. Three predictive criteria, AUC, AARD, and MRD, were considered. When estimating the TPR and FPR, we adopted estimators proposed by Zheng et al. (2013).Predictive results of JM2 were also compared to the other commonly used models, including the Cox model with baseline systolic and diastolic BP measures as fixed covariates, the Cox model with observed longitudinal systolic and diastolic BP measures as time-dependent covariates, JM1 model with only longitudinal

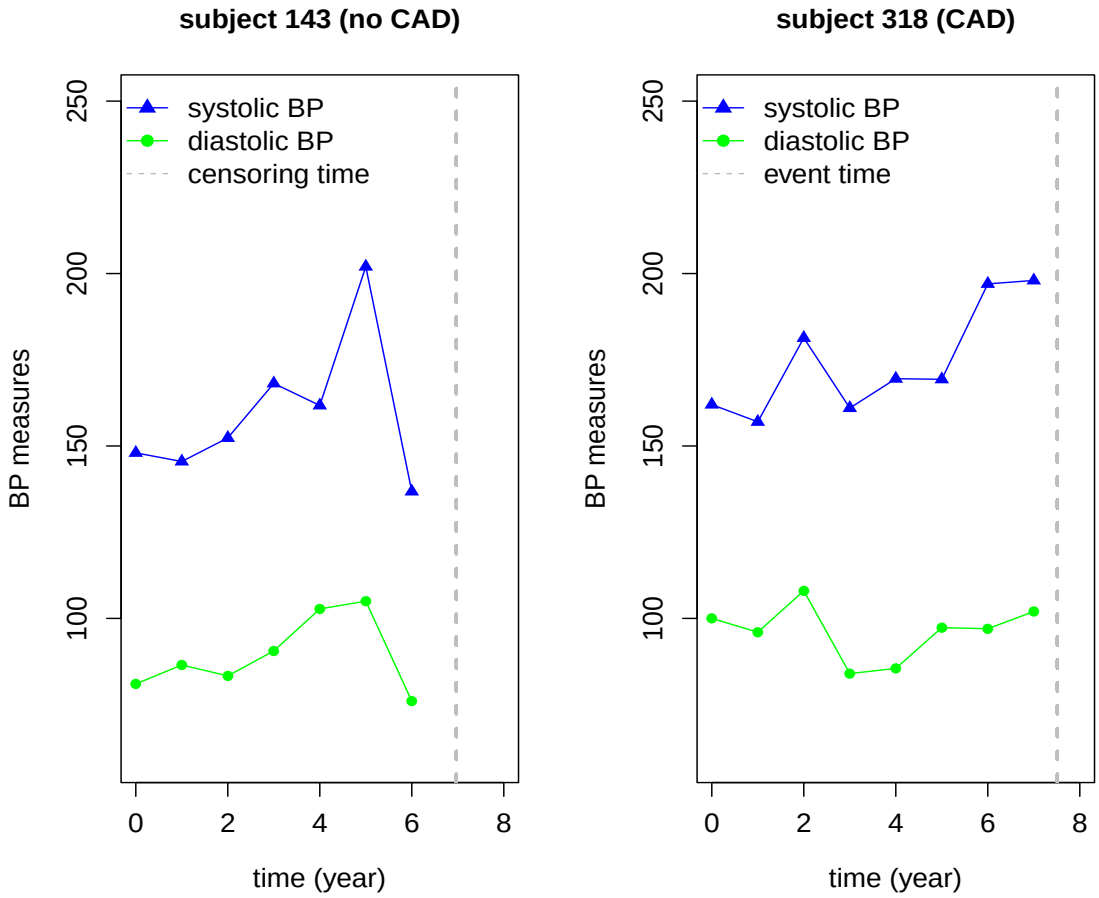

Figure 4.1: Observed longitudinal systolic and diastolic BP measures over time for subject 143 and 318. The blue solid line and triangles denotes the observed systolic BP measures over time. The green solid line and dots depict the observed the diastolic BP measures over time.

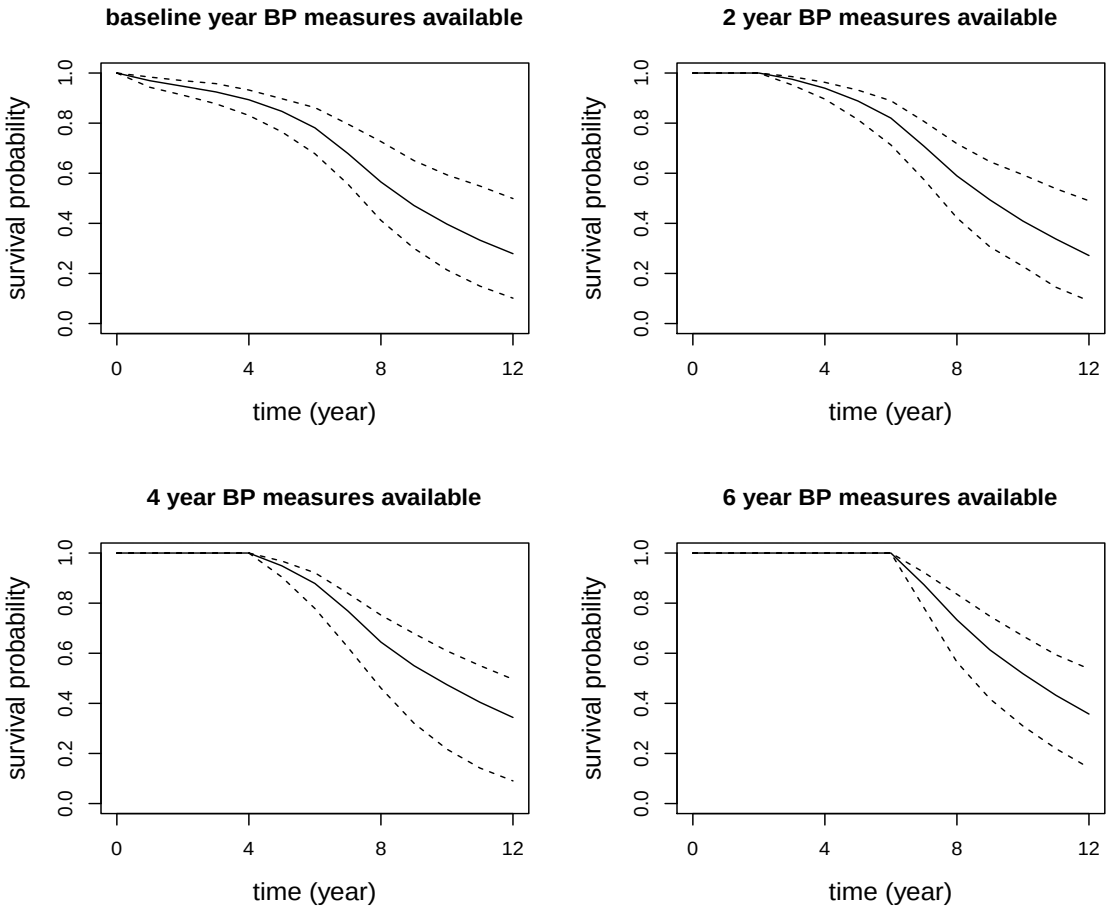

Figure 4.2: Predicted conditional survival probabilities for subject 143. The solid line denotes the median of predicted conditional survival probabilities over the 200 MC samples. The two dashed lines represent the 95% point-wise confidence intervals.

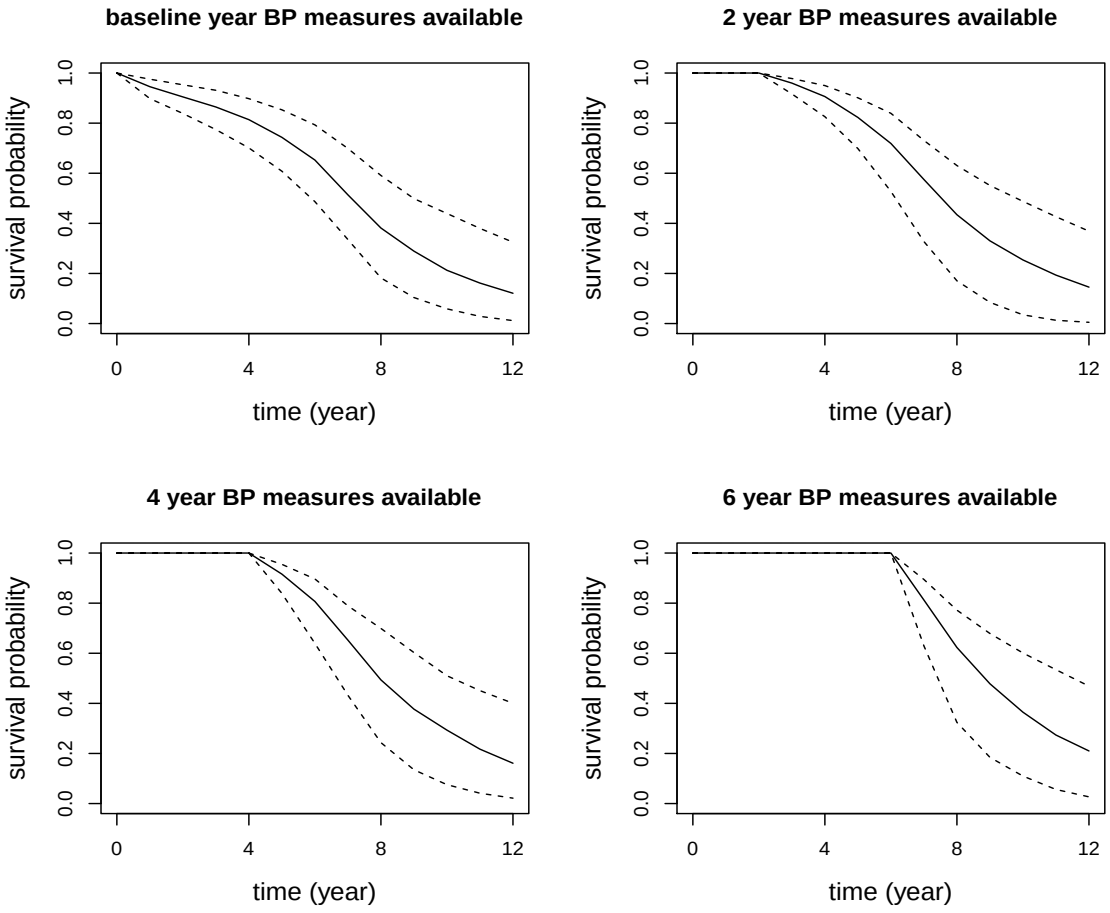

Figure 4.3: Conditional survival probability predictions for subject 318. The solid line denotes the median of predicted conditional survival probabilities over the 200 MC samples. The two dashed lines represent the 95% point-wise confidence intervals.

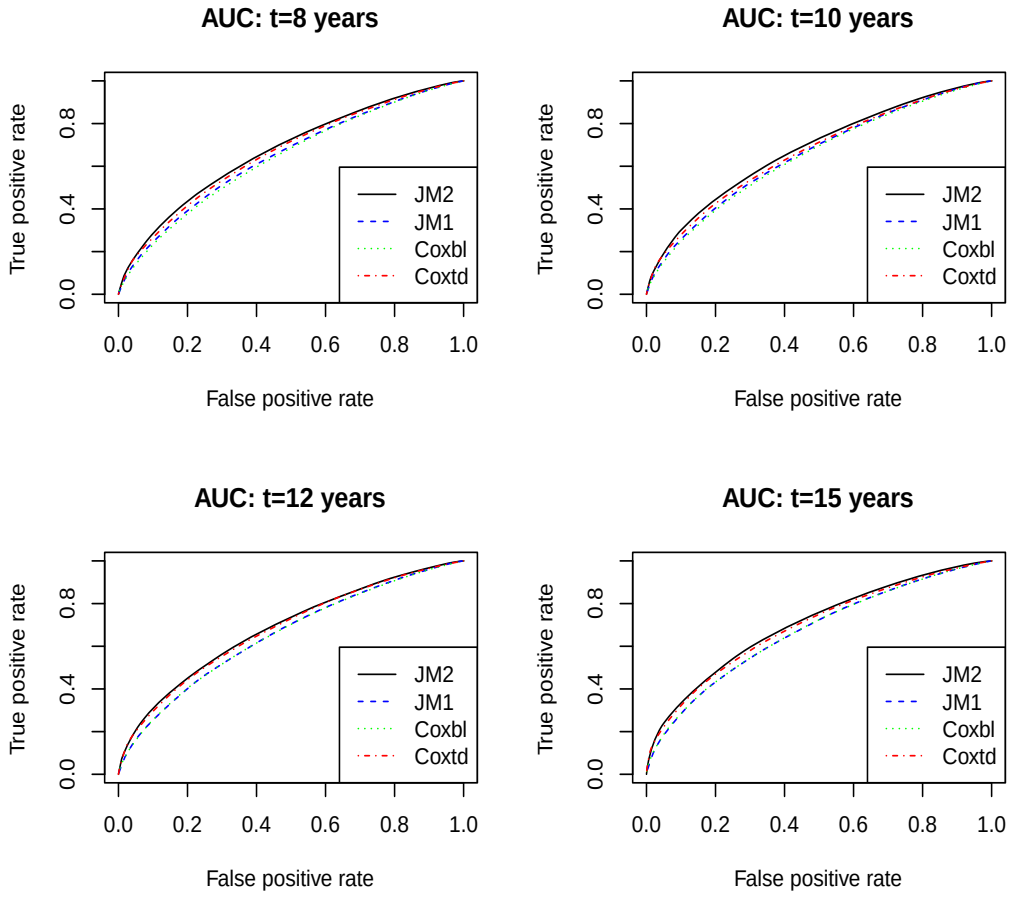

Figure 4.4: Time-dependent ROC curves for different models at different time points.

systolic BP, and JM1 with only longitudinal diastolic BP. Prediction performance comparing different models and criteria at various time points were presented in Table 4.14. Time-dependent ROC curves at different time points and various models were illustrated in Figure 4.4. From these results, it is clear to observe that JM2 model has the best predictive performance: AUC, AARD and MRD of model JM2 are higher than those of all the other models. Particularly, model JM2 has better prediction performance than the two JM1 models indicating that the joint model incorporating both longitudinal systolic and diastolic BP measures can enhance the predictive ability.

Table 4.12: Parameter estimates, standard errors and 95%CI using the training data set. α_1 and α_2 are the association estimates between the risk of CAD and current value of systolic and diastolic BP at event time point, respectively. λ_i $i = 1, ..., 7$ denote the baseline hazards of the 7 piecewise constant intervals.

Parameter	Estimate	StdErr	lower 95%CI	upper 95%CI
Longitudinal Systolic BP				
Intercept	135.20	0.96	133.31	137.08
time	0.33	0.18	-0.03	0.69
time^2	-0.04	0.01	-0.07	-0.02
Age	-0.04	0.06	-0.17	0.08
Race	5.33	0.98	3.41	7.25
$\log(\sigma_1)$	2.49	0.01	2.46	2.52
Longitudinal Diastolic BP				
Intercept	79.19	0.40	78.41	79.97
time	-1.64	0.11	-1.85	-1.43
time^2	0.05	0.01	0.04	0.07
Age	-0.13	0.03	-0.18	-0.08
Race	3.18	0.37	2.45	3.90
$\log(\sigma_1)$	1.94	0.01	1.91	1.97
Time-to-CAD				
Age	0.06	0.01	0.04	0.08
Smoking History	0.36	0.18	0.01	0.71
Race	-0.53	0.19	-0.90	-0.17
Diabetes	0.06	0.17	-0.27	0.40
α_1	0.021	0.010	0.001	0.041
α_2	0.018	0.017	-0.015	0.050
$\log(\lambda_1)$	-8.41	1.00	-10.38	-6.45
$\log(\lambda_2)$	-8.79	1.00	-10.75	-6.83
$\log(\lambda_3)$	-8.43	1.00	-10.38	-6.47
$\log(\lambda_4)$	-7.70	0.98	-9.62	-5.79
$\log(\lambda_5)$	-6.96	0.96	-8.85	-5.07
$\log(\lambda_6)$	-7.09	0.95	-8.96	-5.22
$\log(\lambda_7)$	-6.51	0.92	-8.32	-4.70

Table 4.13: Conditional survival probability predictions for subject 143 and 318. For the MC simulation approach, the median of predictions over 200 MC samples is used as the predicted conditional survival probability. The 2.5% and 97.5% bounds over the 200 MC samples are also presented.

t (year)	Δt (year)	MC simulation approach		
		Median	2.5%CI	97.5%CI
Subject 143				
0	2	0.947	0.912	0.970
	4	0.893	0.831	0.932
	6	0.781	0.678	0.862
2	2	0.939	0.896	0.963
	4	0.820	0.713	0.889
	6	0.589	0.422	0.718
4	2	0.879	0.779	0.921
	4	0.645	0.461	0.752
	6	0.475	0.217	0.609
6	2	0.733	0.565	0.836
	4	0.518	0.310	0.669
	6	0.357	0.143	0.538
Subject 318				
0	2	0.905	0.839	0.953
	4	0.814	0.701	0.898
	6	0.653	0.487	0.793
2	2	0.906	0.826	0.950
	4	0.719	0.527	0.839
	6	0.434	0.171	0.631
4	2	0.807	0.639	0.897
	4	0.494	0.243	0.699
	6	0.293	0.075	0.510
6	2	0.623	0.324	0.772
	4	0.365	0.110	0.601
	6	0.210	0.027	0.469

Table 4.14: Data application results for comparing predictive accuracy criteria of different models.

t(year)	Cox Baseline	Cox Time-dependent	JM1 (sys BP)	JM1 (dias BP)	JM2
			AUC		
8	0.637	0.662	0.646	0.656	0.673
10	0.647	0.664	0.654	0.659	0.678
12	0.652	0.678	0.653	0.674	0.683
15	0.670	0.694	0.670	0.679	0.703
			AARD		
8	0.197	0.234	0.213	0.221	0.249
10	0.210	0.234	0.221	0.222	0.256
12	0.218	0.253	0.217	0.244	0.264
15	0.245	0.280	0.245	0.252	0.296
			MRD		
8	0.052	0.072	0.057	0.066	0.083
10	0.067	0.087	0.073	0.077	0.099
12	0.073	0.102	0.074	0.097	0.108
15	0.085	0.114	0.088	0.099	0.127

4.7 Conclusion

In this chapter we extended the empirical Bayes and the MC simulation approaches (Rizopoulos, 2011) for conditional survival probability prediction in joint models of multiple longitudinal biomarkers and time-to-event outcome. The simulation studies showed that the MC simulation approach induced less bias than the empirical Bayes approach under the large variance scenario. The predicted conditional survival probabilities can be dynamically updated as more longitudinal biomarker measures are collected, and it can help clinicians provide better medical cares for patients.

We also evaluated the predictive accuracy of joint models using AUC, AARD and MRD. Extensive simulations were conducted to assess the performance of proposed methodology. Simulation results reflected that AARD and MRD are more sensitive than AUC in quantifying the improved predictive ability by adding new longitudinal biomarker in the joint models. In addition, simulation results implied that joint models have better predictive accuracy than the commonly used Cox models. In simulations, it is also observed that the prediction accuracy of random effect has some sort of impact on the predictive accuracy of joint models. In the future, we plan to study for new approaches to improve accuracy of random effect prediction in the joint modeling framework.

In clinical trials and observational studies, it's common to see longitudinal biomarkers are also collected besides of the primary time-to-event end point. These longitudinal biomarkers could be very useful indicators for the disease progression since the underling trajectories of the longitudinal biomarkers can be very informative. For example, the CD4 cell count measures are often measured in the study of human immunodeficiency virus (HIV)- the lower CD4 count indicates the higher risk of HIV; in studies of prostate cancer, the prostate specific antigen (PSA) level is also frequently measured since high PSA is a strong risk factor of prostate cancer; in cardiovascular disease studies, the data of common

risk factors, systolic and diastolic blood pressures, lipids and cholesterol levels, are also routinely collected. A natural question will be whether including the longitudinal biomarker trajectory in the model can improve the predictive ability of the model as well as how much improvement can the new longitudinal biomarker provide. The existing methodology of joint models prediction can only be used to assess the predictive performance of one single longitudinal biomarker. In reality, multiple longitudinal biomarkers could be associated with the disease risk. Often times, the biomarker assessment could be highly costly. So it would be very useful and practical to decide how much prediction improvement can be gained by adding another new longitudinal biomarker in to the existing model. Our proposed methodology of prediction for joint model with multiple longitudinal outcomes are well established to answer such questions. In our data application, we demonstrated that the joint models with both systolic and diastolic BP have better prediction performance than joint models with only one type of BP measures. The methodology can be applied to many other clinical or epidemiology studies.

4.8 Acknowledgement

The research is supported by National Institutes of Health (NIH) Grants R01 AG019181, R24 MH080827, and P30 AG10133.

Chapter 5

Conclusion

In this thesis we have studied several topics related to joint models for longitudinal and survival data analysis. The proposed methodologies are applicable to many medical research areas.

In Chapter 2, we developed joint modeling frameworks of bivariate longitudinal outcomes under three different bivariate random change point models: the bivariate random broken-stick model, the bivariate random Bacon-Watts model, and the bivariate random smooth polynomial model. The Bayesian method was used for model fitting using the BRugs package in **R**. The proposed bivariate random change point models not only estimate the change points of bivariate longitudinal outcomes, but also investigate the correlation between the change points. This methodology is useful for disease prognosis using biomarkers in medical science, and it also possesses flexibility in model fitting. Although we have focused on investigating the correlation between the change points, one can readily extend to more complex models by specifying and estimating other correlation parameters such as correlation between the two slopes before the change point as well as the two slopes after the change point in the bivariate model. The extension to multivariate change point models for multiple longitudinal outcomes is also applicable.

Joint models for multiple longitudinal biomarkers and time-to-event outcome were investigated in Chapter 3. We developed a maximum-likelihood method using the EM algorithm for parameter estimation of joint models. We conducted a series of simulations to assess the performance of EM algorithm in parameter estimation of joint models. The simulations compared the existing two-stage method to the EM algorithm, and confirmed that

the EM algorithm can provide more accurate and efficient parameter estimates than the two-stage method. The methodology of joint models for multiple longitudinal processes and time-to-event outcome are applicable to many clinical trials and observational studies where the association between multiple longitudinal biomarkers and time-to-event outcome is of interest. In this work, we focused on joint models with normally distributed longitudinal outcomes, however, it is straightforward to extend to joint models with various distributed longitudinal outcomes, including binary, poisson outcome and so on. The proposed EM algorithm can be used to estimate parameters of joint models with these mixed types of longitudinal outcomes. On the other hand, we can extend the standard Cox PH models to compete-risk model or semi-compete-risk model to take into account of non-uniformly distributed time-to-event outcome.

Last, in Chapter 4 we continued the joint modeling framework in the previous chapter and focused on the predictive accuracy of joint models. We assessed the MC simulation approach (Rizopoulos, 2011) for the conditional survival probability prediction in joint models for multiple longitudinal biomarkers and a time-to-event outcome. The predicted conditional survival probability can be dynamically updated as more longitudinal biomarker measures are collected, and it can help clinicians provide better medical cares for patients. We also evaluated the predictive accuracy of the joint models using AUC, AARD and MRD. We conducted extensive simulations to assess the performance of proposed methodology. Simulation results reflect that AARD and MRD are more sensitive than AUC in quantifying the improved predictive ability introduced by adding new biomarkers in the joint models. The methodology of the predictive accuracy of the joint models are applicable to many clinical trials and observational studies.

Joint models of multiple longitudinal outcomes and time-to-event outcome can find many applications in observational studies or clinical trial settings. Currently many cohort

studies still rely on baseline measures when determining the associations between putative risk factors and time to event. Such approaches make the implicit assumption that the exposure measures stay constant over the observation period. However, in many medical research areas, such an assumption is unlikely to hold. In aging research, for example, many biomarkers were found to change with time either as part of the aging process, disease progression or in response to treatment.

The use of time-dependent Cox models, although no longer requiring the assumption of constant exposure, has its own limitations, as we described in details in Chapter 3. The first is that longitudinal exposure levels were assumed to be measured without error; the second is that trends in the longitudinal processes cannot be used to determine their association with time to event. The joint modeling framework we used in this research is capable of including various aspects of the longitudinal outcomes such as current values, rate of changes or cumulative exposures as part of the model framework so that we can identify the attributes in the longitudinal outcomes that are most significantly related to the survival outcome.

The joint modeling framework offers a powerful tool for data collected in routine clinical practice where biomarkers such as blood pressure, lipids, glucose level and other markers are collected as part of physical examination conducted at regular intervals. The use of joint modeling approach can offer prospective prediction of patients' risk for various diseases based on the collective longitudinal measures of these marker values and give clinicians a better tool to synthesize these repeatedly measured marker values. In health care system with comprehensive electronic medical record data, the modeling and prediction approach can be incorporated into the system to offer real-time modeling and prediction, providing on-going decision support to clinicians and leading ultimately to better patients' care and health outcomes.

BIBLIOGRAPHY

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika* **52**, 317–332.

Albert, P. S. and Shih, J. H. (2010). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics* **66**, 983–987.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

Andersen, P. K. and Gill, R. D. (1982). Cox Regression Model for Counting Processes: A large Sample Study. *Annals of Statistics* **10**, 1100–1120.

Anderson, K. M., Odell, P. M., Wilson, P. W. and Kannel, W. B. (1991). Cardiovascular disease risk profiles. *American Heart Journal* **121**, 293–298.

Antolini, F., Boracchi, P. and Biganzoli, E. (2005). A time-dependent discrimination index for survival data. *Statistics in Medicine* **24**, 3927–3944.

Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525–534.

Brown, E. R. and Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59**, 221–228.

Brown, E. R., Ibrahim, J. G. and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61**, 64–73.

Buchman, A. S., Wilson, R. S., Bienias, J. L., Shah, R. C., Evans, D. A. and Bennett, D. A. (2005). Change in body mass index and risk of incident Alzheimer disease. *Neurology* **65**, 892–897.

Callahan, C. M., Hendrie, H. C., Dittus, R. S., Brater, D. C., Hui, S. L. and Tierney, W. M. (1994). Improving treatment of late life depression in primary care: a randomized clinical trial. *Journal of the American Geriatrics Society* **42**, 839–846.

Callahan, C. M., Hui, S. L., Nienaber, N. A., Musick, B. S. and Tierney, W. M. (1994). Longitudinal study of depression and health services use among elderly primary care patients. *Journal of the American Geriatrics Society* **42**, 833–838.

Chambless, L. E., Cummiskey, C. P. and Cui, G. (2011). Several methods to assess improvement in risk prediction models: extension to survival analysis. *Statistics in Medicine* **30(1)**, 22–38.

Chambless, L. E. and Diao, G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine* **25**, 3474–3486.

Chen, M. H., Shao, Q. M. and Ibrahim, J. G. (2000). *Monte Carlo methods in Bayesian computation*. Springer-Verlag, NewYork.

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* **115**, 928–935.

Couvreur, C. The EM algorithm: A guided tour. *In Proc. 2d IEEE European Workshop on Computationaly Intensive Methods in Control and Signal Processing (CMP' 96), pages 115 to 120, Pragues, Czech Republik, August 1996*.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B* **34(2)**, 187–220.

Dafni, U. G. and Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome whenmeasured with error. *Biometrics* **54**, 1445–1462.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood estimation from imcplete data via EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)* **39**, 1–38.

Dominicus, A., Ripatti, S., Pedersen, N. L. and Palmgren, J. (2008). A random change point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine* **27**, 5786–5798.

Elashoff, R., Li, G. and Li, N. (2006). Joint Models for Multivariate Longitudinal and Multivariate Survival Data. *Biometrics* **62**, 432–445.

Elashoff, R., Li, G. and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64**, 762–771.

Faucett, C. J. and Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* **15**, 1663–1685.

Fieuws, S., Verbeke, G., Maes, B. and Vanrenterghem, Y. (2008). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics* **9**, 419–431.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley and Sons.

Gao, S., Nguyen, J. T., Hendrie, H. C., Unverzagt, F. W., Hake, A., Smith-Gamble, V. and Hall, K. (2011). Accelerated weight loss and incident dementia in an elderly African-American cohort. *The American Geriatrics Society* **59**, 18–25.

Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). Technical report, Bayesian Statistics 4, Oxford University Press, Oxford.

Gerds, T. A., Cai, T. and Schumacher, M. (2008). The performance of risk prediction models. *Biometrical Journal* **50**, 457–479.

Ghosh, P., Ghosh, K. and Tiwari, R. C. (2010). Joint modeling of longitudinal data and informative dropout time in the presence of multiple changepoints. *Statistic in Medicine* **30**, 611–626.

Ghosh, P. and Vaida, F. (2007). Random change point modelling of HIV immunologic responses. *Statistics in Medicine* **26**, 2074–2087.

Gonen, M. and Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* **92(4)**, 965–970.

Gueorguiva, R. and Sanacora, G. (2006). Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine* **25**, 1307–1322.

Hall, C. B., Ying, J., Kuo, L. and Lipton, R. B. (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics and Data Analysis* **42**, 91–109.

Hall, C. B., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M. and Lipton, R. B. (2001). Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in Medicine* **20**, 3695–3714.

Hall, K. S., Ogunniyi, A. O., Hendrie, H. C., Osuntokun, B. O., Hui, S. L., Musick, B. S., Rodenberg, C. A., Unverzagt, F. W., Gureje, O. and Baiyewu, O. (1996). A cross-cultural community based study of dementias: methods and performance of the survey

instrument: Indianapolis, U.S.A. and Ibadan, Nigeria. *International Journal of Methods in Psychiatric Research* **6**, 129–142.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143(1)**, 29–36.

Harrell, F. E. (2001). *Regression Modeling Strategies*. New York: Springer.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. and Rosati, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association* **247**, 2543–2546.

Harrell, F. E., Lee, K. L. and Mark, D. B. (1996). Tutorial in Biostatistics: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* **15**, 361–387.

He, B. and Luo, S. (2013). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research*.

Heagerty, P. J., Lumley, T. and Pepe, M. S. (2000). Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* **56(2)**, 337–344.

Heagerty, P. J. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics* **4**, 465–480.

Hendrie, H. C., Ogunniyi, A., Hall, K. S., Baiyewu, O., Unverzagt, F. W., Gureje, O., Gao, S., Evans, R. M., Ogunseyinde, A. O., Adeyinka, A. O., Musick, B. S. and Hui, S. L. (2001). Incidence of dementia and Alzheimer disease in 2 communities: Yoruba residing

in Ibadan, Nigeria, and African Americans residing in Indianapolis, Indiana. *the Journal of the American Medical Association* **285**, 739–747.

Hendrie, H. C., Osuntokun, B. O., Hall, K. S., Ogunniyi, A. O., Hui, S. L., Unverzagt, F. W., Gureje, O., Rodenberg, C. A., Baiyewu, O., Musick, B. S., Adeyinka, A., Farlow, M. R., Oluwole, S., Class, C. A., Komolafe, O., Brashear, A. and Burdine, V. (1995). Prevalence of Alzheimer's disease and dementia in two communities: Nigerian Africans and African Americans. *The American Journal of Psychiatry* **152**, 1485–1492.

Hochman, J., Tamis, J. E. and Thompson, T. D. (1999). Sex, clinical presentation, and outcome in patients with acute coronary syndromes. *New England Journal of Medicine* **341**, 226–232.

Huang, Y., Dagne, G. and Wu, L. (2011). Bayesian inference on joint models of HIV dynamics for time-toevent and longitudinal data with skewness and covariate measurement errors. *Statistics in Medicine* **30**, 2930–2946.

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C. and Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology* **9**, 119–128.

Jacqmin-Gadda, H., Commenges, D. and Dartigues, J.-F. (2006). Random change point model for joint modeling of cognitive decline and dementiae. *Biometrics* **62**, 254–260.

Janes, H., Pepe, M. S. and Gu, W. (2008). Assessing the value of risk predictions by using risk stratification tables. *Annals of Internal Medicine* **149**, 751–760.

Kiuchi, A. S., Hartigan, J. A., Holford, T. R., Rubinstein, P. and Stevens, C. E. (1995). Change points in the series of T4 counts prior to AIDS. *Biometrics* **51**, 236–248.

Lange, N., Carlin, B. P. and Gelfan, A. E. (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *Journal of the American Statistical Association* **87**, 615–626.

Ligges, U., Thomas, A., Spiegelhalter, D., Best, N., Lunn, D., Rice, K. and Sturtz, S. (2009). BRugs 0.5: OpenBUGS and Its R/S-PLUS Interface BRugs.

Little, R. J. (1993). Pattern-Mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88(421)**, 125–134.

Little, R. J. and Rubin, D. B. (2001). *Statistical Analysis with Missing Data* (second ed.). New York: John Wiley.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

McCulloch, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research* **17**, 53–73.

Moons, K. G. and Harrell, F. E. (2003). Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. *Academic Radiology* **10**, 670–672.

Njagi, E. N., Rizopoulos, D., Molenberghs, G., Dendale, P. and Willekens, K. (2013). A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients. *Statistical Modelling* **13**, 179–198.

Njeru Njagi, E., Molenberghs, G., Verbeke, G., Kenward, M. G., Dendale, P. and Willekens, K. (2013). A flexible joint-modelling framework for longitudinal and time-to-event data with overdispersion. *Statistical Methods in Medical Research*.

Pencina, M. J. and Agostino, S. R. B. D. (2004). Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* **23**, 2109–2123.

Pencina, M. J., D'Agostino, R. B., D'Agostino, S. R. B. and Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine* **27**, 157–172.

Pencina, M. J., D'Agostino, R. B. and Steyerberg, E. W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine* **30(1)**, 11–21.

Pencina, M. J., D′Agostino, S. R. B. and Song, L. (2012). Quantifying discrimination of Framingham risk functions with different survival C statistics. *Statistics in Medicine* **31(15)**, 1543–1553.

Pepe, M. S., Feng, Z. and Gu, J. (2008). Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine. *Statistics in medicine* **27**, 173–181.

Pepe, M. S., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. M. and Zheng, Y. (2008). Integrating the predictiveness of a marker with its performance as a classifier. *American Journal of Epidemiology* **167**, 362–368.

Pepe, M. S. and Janes, H. (2012). Methods for Evaluating Prediction Performance of Biomarkers and Tests. *(University of Washington Biostatistics Working Paper Series, working paper 384).*

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (2007). *Numerical Recipes: The Art of Scientific Computing* (Third ed.). New York: Cambridge University Press,.

Proust-Lima, C., Séne, M., Taylor, J. M. and Jacqmin-Gadda, H. (2012). Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research*.

Proust-Lima, C. and Taylor, J. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: A joint modeling approach. *Biostatistics* **10**, 535–549.

Qiu, F., Stein, C. M. and Elston, R. C. (2013). Joint modeling of longitudinal data and discrete-time survival outcome. *Statistical Methods in Medical Research*.

R Development Core Team (2007). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing: Vienna, Austria*.

Rizopoulos, D. (2010). JM : An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* **35**, 1–33.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67**, 819–829.

Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics and Data Analysis* **56**, 491–501.

Rizopoulos, D. (2012b). *Joint Models for Longitudinal and Time-to-Event Data, with Applications in R.* Chapman and Hall/CRC Biostatistics Series.

Rizopoulos, D. and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* **30**, 1366–1380.

Rizopoulos, D., Hatfield, L. A., Carlin, B. P. and Takkenberg, J. J. (2013). Combining Dynamic Predictions from Joint Models for Longitudinal and Time-to-Event Data using Bayesian Model Averaging.

Rothwell, P. M., Howard, S. C., Dolan, E., O'Brien, E., Dobson, J. E., Dahlöf, B., Sever, P. S. and Poulter, N. R. (2010). Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. *Lancet* **375**, 895–905.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics* **6(2)**, 461–464.

Self and Pawitan (1992). Modeling a marker of disease progression and onset of disease. *AIDS Epidemiology: Methodological Issues*.

Song, X., Davidian, M. and Tsiatis, A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* **3**, 511–528.

Sousa, I. (2011). A review of joint modeling of longitudinal measurements and time-to-event. *REVSTAT* **9**, 57–81.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. V. D. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639.

Stamler, J., Stamler, R. and Neaton, J. D. (1993). Blood pressure, systolic and diastolic, and cardiovascular risks. US population data. *Archives of Internal Medicine* **153(5)**, 598–615.

Sweeting, M. J. and Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal* **53**, 750–763.

Taylor, J., Yu, M. and Sandler, H. (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of Clinical Oncology* **23**, 816–825.

Taylor, J. M. G., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T. and Sandler, H. (2013). Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models. *Biometrics* **69**, 206–213.

Tseng, Y. K., Hsieh, F. and Wang, J. L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika* **92**, 587–603.

Tsiatis, A. A. and Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica* **14**, 793–818.

Tsiatis, A. A., DeGruttola, V. and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90**, 27–37.

Uno, H., Cai, T., Pencina, M. J., Agostino, R. D. and Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine* **30**, 1105–1117.

Vaccarino, V., Parsons, L., Every, N. R., Barron, H. V. and Krumholz, H. M. (1999). Sex-based differences in early mortality after myocardial infarction. National Registry of Myocardial Infarction 2 Participants. *New England Journal of Medicine* **341**, 217–225.

van den Hout, A., Muniz-Terrera, G. and Matthews, F. E. (2010). Smooth random change point models. *Statistics in Medicine* **30**, 599–610.

Wilson, P. W. F., Agostino, R. B. D., Levy, D., Belanger, A. M., Silbershatz, H. and Kannel, W. B. (1998). Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* **97**, 1837–1847.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society Series C* **50**, 375–387.

Yang, L. and Gao, S. (2012). Bivariate random change point models for longitudinal outcomes. *Statistics in Medicine* **32(6)**, 1038–1053.

Yates, J. F. (1982). External correspondence: decomposition of the mean probability score. *Organ Behav Hum Perform* **30**, 132–156.

Ye, W., Lin, X. and Taylor, J. (2006). A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and its Interface* **1**, 33–45.

Yu, M., Law, N., Taylor, J. and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* **14**, 835–862.

Yu, M., Taylor, J. and Sandler, H. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association* **103**, 178–187.

Zheng, Y., Cai, T. and Pepe, M. S. (2013). Adopting nested case–control quota sampling designs for the evaluation of risk markers. *Lifetime data analysis*, 1–21.

Zheng, Y. and Heagerty, P. J. (2007). Prospective Accuracy for Longitudinal Markers. *Biometrics* **63**, 332–341.

CURRICULUM VITAE

Lili Yang

EDUCATION

- Ph.D. in Biostatistics, Indiana University, Indianapolis, IN, 2013 (minor in Epidemiology)
- M.S. in Applied Mathematics, NC State University, Raleigh, NC, 8/2009
- M.S. in Computational Mathematics, Nanjing Normal University, Nanjing, China, 7/2006
- B.S. in Applied Mathematics, Nanjing Normal University, Nanjing, China, 7/2003

WORKING EXPERIENCE

- Summer Intern, Eli Lilly and Company, Indianapolis, IN, 5/2012 - 8/2012
- Research Assistant, Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, 8/2010 - 12/2013
- Teaching Assistant, Department of Mathematics, North Carolina State University, Raleigh, NC, 1/2007 - 5/2009
- Teaching Aassistant, Department of Mathematics, Nanjing Normal University, Nanjing, China, 9/2005 - 6/2006

SELECT PUBLICATIONS

- Yang, L., Menggang Yu and Gao, S. Dynamic predictions and prospective accuracy in joint models for multiple longitudinal processes and time-to-event outcome. In preparation.

- Yang, L., Menggang Yu and Gao, S. Joint models for multiple longitudinal processes and time-to-event outcome. In preparation.
- Yang, L. and Gao, S. (2012). Bivariate Random Change Point Models for Longitudinal Outcomes. Statistics in Medicine, 32, Issue 6.
- Yang, L., Unverzagt, F. W., Jin, Y., Hendrie, H. C., Liang, C., Hall, K. S., Cao, J., Ma, F., Murrell, J. R., Cheng, Y., Li, P., Bian, J., Gao, S. (2012). Normative Data for Neuropsychological Tests in a Rural Elderly Chinese Cohort. The Clinical Neuropsychologist 26, 641-653.
- Gao, S., Hendrie, H. C., Yang, L., Stump, T. E. et al. (2012). Increase in blood pressure precedes clinical depression in elderly primary care patients. Accepted by International Journal of Geriatric Psychiatry.
- Wang, H., Jin, Y., Hendrie, H. C., Liang, C., Yang, L., et al. (2012). Late life leisure activities and risk of cognitive decline. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 68, 205-213.
- Kuruppu, D., Hendrie, H. C., Yang, L., Gao, S. (2012). Selenium Levels and Hypertension in Rural Elderly Chinese. Submitted.
- Kuruppu, D., Hendrie, H. C., Yang, L., Gao, S. (2011). Selenium Levels and Hypertension: A Systematic Review of the Literature. Accepted by Public Health Nutrition.
- Yang, L. and Zhang, Z. (2006). Finite Volume Element Methods for One Dimensional Malaria-Malaria Mosquito Model with Diffusion. Journal of Anshan Normal University 8, Issue 2.